Constraints on Mixing and CP-Violation in the Neutral Charmed Meson System at LHCb

Michael T. Alexander

University of Glasgow
Department of Physics and Astronomy

Submitted in fulfilment of the requirements
for the degree of Doctor of Philosophy

May 2012

Abstract

This thesis presents measurements of the charm sector mixing and CP-violation parameters y_{CP} and A_Γ, made using data collected in 2010 by the LHCb experiment at the LHC at $\sqrt{s} = 7$ TeV. y_{CP} is defined as the difference from unity of the ratio of the effective lifetime of the D^0 meson decaying to a CP-undefined final state to its lifetime when decaying to a CP-eigenstate. A_Γ is the CP-asymmetry of the effective lifetimes of the D^0 and $\overline{\mathrm{D}}^0$ when decaying to a CP-eigenstate. In the absence of CP-violation y_{CP} will be consistent with the mixing parameter y, and A_Γ will be consistent with zero.

CP-violation in the charm sector is predicted to be very small in the SM, though first evidence for direct CP-violation in D^0 decays has recently been observed by LHCb. Observation of significantly more CP-violation than is allowed in the SM would be a strong indication of new physics. The current world best measurements of y_{CP} and A_Γ show no evidence of CP-violation.

The methods used to measure the effective lifetime of the D^0 are presented, together with a detailed study of the impact parameter resolutions achieved by Vertex Locator (VELO) sub-detector. A resolution of $< 36\,\mu\mathrm{m}$ on the x and y components of impact parameter measurements is measured for particles with $p_T > 1\,\mathrm{GeV}$. The final dataset on which y_{CP} and A_Γ are measured comprises $28.0 \pm 2.8\,\mathrm{pb}^{-1}$, from which 226,110 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^\mp\pi^\pm$ and 30,481 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^+\mathrm{K}^-$ candidates are selected. The dominant sources of systematic uncertainty arise from combinatorial backgrounds and D^0 produced in decays of B mesons. The final results are found to be

$$\begin{aligned} A_\Gamma &= (-5.9 \pm 5.9\,(\mathrm{stat.}) \pm 2.1\,(\mathrm{syst.})) \times 10^{-3}, \\ y_{CP} &= (5.5 \pm 6.3\,(\mathrm{stat.}) \pm 4.1\,(\mathrm{syst.})) \times 10^{-3}. \end{aligned}$$

y_{CP} is consistent with the world average of y and with zero, and A_Γ is consistent with zero. Thus, these results show no evidence for CP-violation or mixing in the D^0 system.

Acknowledgements

The writing of this thesis and the years of study that preceded it has been an epic adventure that will stay with me forever. It has given me many opportunities and challenges, both amazing and trying, that have shaped the course of my life in ways I couldn't have conceived four years ago. Consequently, there are many people to whom I owe thanks for the parts they have played in my journey.

Firstly, there is the University of Glasgow, my supervisors, Chris Parkes and Paul Soler, and our Professor, Tony Doyle, for taking me on in the first place. Given my astounding ignorance of the field back when I interviewed for the PhD position I've never quite understood how I managed to get the offer (I think it must have been through sheer enthusiasm), but I will be eternally grateful for having been given it. I must also thank the STFC for funding me throughout, particularly for affording me my 18 month stay in Geneva.

Of course, my supervisors did much more than simply take me on. They have served as my guides through the bizarre world of physics and academia, ensuring that I stayed on track, and that I took advantage of all the many opportunities that were available to me. In addition, I am very grateful for their careful proof reading of this thesis, and the vast amount of advice given on how to improve it. My many collaborators have also been invaluable in getting our research to the stage at which it is today, giving me plenty about which to write. So my great thanks go to Vava (particularly for the early days), Marco (for having the ideas in the first place), and Silvia (for working with me through the day to day slog, and particularly for bringing the studies of systematic uncertainties together). Thanks also to Lars and Paul Sail (for helping us show that our methods actually work), and to Paula Collins (for all the encouragement and advice given in my studies of the VELO). In addition, I should thank Valerie, our group secretary, and the staff at the UK liaison office at CERN, for making my many trips to and from Geneva and elsewhere much less stressful than they could have been.

Finally, on the academic side of things, I must thank my examiners, Dr Rick St. Denis and Dr Todd Huffman, without whom I wouldn't be in a situation to thank anyone. Thanks to them for taking the time to read this thesis in detail, for grilling me quite so thoroughly on it in my viva, and for making my viva not quite as painful as it could have been.

For support far beyond the academic realm I cannot thank my fiancee Emma enough. She has stuck with me throughout, listening to my moaning and sharing in the occasional successes, and remained supportive even during my long stay in Geneva. Even more so, she has constantly reminded me of what an amazing opportunity I have been given, and kept me from ever taking it for granted. She has also kept me motivated and kept me going at times when my resolve was failing. This thesis is as much her achievement as it is mine, and I look

forward to sharing many more successes with her in our life together.

My other family members have also been a never ending source of support. My parents have always let me know that they are proud of me and listened interestedly to the latest developments in my work. My soon-to-be parents in-law have also been endlessly supportive, particularly for the long period during the writing of this thesis for which I was effectively living with them.

Others to whom I owe thanks include my many friends, for helping keep me sane and generally have good times in amongst all the work. Carson in particular for giving me his thesis template to get me started! The guys from Glasgow, Gen and Aurelo, Alistair and Jenny, Cristy and Nacho, and the many friends I made while staying in Geneva, particularly the guys from Imperial College, Arlo and Cat, Alex, Mike, and Paul, and also Ravi, where ever he may be. Hopefully the good times are far from over!

Finally I'd like to say good luck to the current batch of PhD students at Glasgow, particularly Paul, Mike and Nick, all of whom are starting to write their own theses, and also to Daniel, and Stephen, and John, particularly as he is taking over the impact parameter studies. I hope you all have as much fun and get as much out of your PhDs as I have.

Declaration

The research results presented in this thesis are the product of my own work. Appropriate references are provided when results of third parties are mentioned. The research presented here was not submitted for another degree in any other department or university.

Michael T. Alexander

Preface

This thesis presents measurements of the charm sector mixing and CP-violation parameters y_{CP} and A_Γ, made using data collected in 2010 by the LHCb experiment at the LHC at $\sqrt{s} = 7$ TeV. The Vertex Locator (VELO) sub-detector of LHCb is particularly important in performing these measurements. Hence, the performance of the VELO, with respect to the resolutions it achieves on impact parameter (IP) measurements, is also studied in detail.

Chapter 1 presents a review of the Standard Model (SM) of particle physics. The general theory is discussed, with particular emphasis placed on mixing and CP-violation in the charm quark sector. The parameters y_{CP} and A_Γ, and their dependence on the level of mixing and indirect CP-violation in the D^0 system, are then detailed. These are defined as

$$y_{CP} = \frac{\tau_{\text{eff}}(\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^\mp\pi^\pm)}{\tau_{\text{eff}}(\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^+\mathrm{K}^-)} - 1, \quad \text{and}$$
$$A_\Gamma = \frac{\tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-) - \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)}{\tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-) + \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)}.$$

In the absence of CP-violation y_{CP} will be consistent with the mixing parameter y, while A_Γ will be consistent with zero. y_{CP} is thus primarily a measurement of mixing in the D^0 system, and requires an independent measurement of y to check for CP-violation.

CP-violation in decays involving c quarks is predicted to be $\mathcal{O}(10^{-3})$ or less in the SM. Observation of significantly more CP-violation would be indicative of new physics. It is thus very exciting that direct CP-violation in the D^0 system has recently been observed at $\mathcal{O}(10^{-3})$ by LHCb [1], which is around the upper limit allowed in the SM. The BABAR experiment at SLAC has made the current best measurement of $y_{CP} = (11.6 \pm 2.2\,(\text{stat.}) \pm 1.8\,(\text{syst.})) \times 10^{-3}$ [2]. This excludes the no mixing hypothesis at 4.1σ, and is consistent with the world average of $y = (8.0 \pm 1.3) \times 10^{-3}$ [3]. The current best measurement of A_Γ comes from the BELLE experiment at KEK, who find $A_\Gamma = (0.1 \pm 3.0\,(\text{stat.}) \pm 1.5\,(\text{syst.})) \times 10^{-3}$ [4]. Thus, no evidence for indirect CP-violation in the D^0 system is currently observed.

The LHC accelerator complex and the LHCb detector are described in chapter 2. The VELO is described in particular detail. It is a silicon strip detector that measures the positions of any p-p collisions and displaced decay vertices in an event. It is thus essential in measuring the lifetime of a long lived particle, like the D^0. The Ring Imaging Cherenkov (RICH) sub-detectors, which provide particle identification, are also detailed. They provide clean separation of πs and Ks, and so help minimise backgrounds from particle mis-identification.

The IP resolutions achieved by the VELO are studied in chapter 3. An excellent resolution on IP_x of $< 36\,\mu\mathrm{m}$ is measured for particles with $p_T > 1\,\mathrm{GeV}$. A mathematical parametrisation of IP resolutions is derived, depending on the single hit resolution, material budget and extrapolation distance to the interaction point. The predictions of this parametrisation are compared to measurements made on 2011 data and data obtained from full Monte Carlo simulation. In general IP resolutions are found to behave as expected according to the parametrisation. A momentum dependent discrepancy in the resolutions measured on real and simulated data is observed. This suggests that the description of the material in the VELO is not entirely accurate in the simulation. However, complementary studies of the distribution of material in the VELO have not been able to confirm or deny this finding.

A method by which one can extract the effective lifetime of a long lived particle from a dataset including signal and backgrounds is presented in chapter 4. Prompt D^0, produced directly at the p-p collision, are used to measure the lifetime. The backgrounds considered thus comprise 'secondary' D^0, produced in decays of B mesons, and combinatorial backgrounds. First, a fit is performed to the distribution of the reconstructed mass of the D^0 to extract the signal yield and distinguish combinatorial backgrounds. For the measurement of A_Γ the flavour of the D^0 at production is tagged using the decay chain $\mathrm{D}^{*\pm} \to \mathrm{D}^0\pi^\pm$, where the charge of the $\pi^\pm$ gives the flavour of the D^0. An additional background enters here when a correctly reconstructed D^0 is combined with a random $\pi^\pm$ to make the $\mathrm{D}^{*\pm}$, so that the D^0 is assigned a random flavour tag. Such candidates are distinguished using a simultaneous fit to the distributions of the reconstructed mass of the D^0 and the difference between that and the reconstructed mass of the $\mathrm{D}^{*\pm}$. Secondary D^0 cannot be distinguished using the mass distributions, but tend to have larger IPs than prompt D^0 at high proper decay times. Hence, they are distinguished using a simultaneous fit to the proper-decay-time and D^0 IP distributions.

The fit to the proper decay time distribution also gives the lifetime of the signal candidates. However, this fit must correct for lifetime biasing selection criteria used in the trigger and offline candidate selections. A data-driven method of performing this correction is also discussed in chapter 4. This involves artificially changing the proper decay time of each D^0 candidate in the dataset and re-evaluating the decision of the candidate selection at each proper decay time. Thus, one obtains the selection efficiency as a function of proper decay time for each candidate. The manner in which the acceptance functions obtained via this method are incorporated into the fit PDFs is also covered. The full fit PDF can be used to distinguish signal and all backgrounds, and correct for lifetime biasing candidate selections.

These methods are applied to $28.0 \pm 2.8\,\mathrm{pb}^{-1}$ of data collected by LHCb in 2010. The results and their statistical uncertainties are presented in chapter 5. Firstly, the trigger and

offline selection criteria applied to the data are discussed. Strong lifetime biasing selection criteria are used to exclude combinatorial backgrounds. The final datasets comprise 286,159 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{\mp}\pi^{\pm}$ and 39,263 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{+}\mathrm{K}^{-}$ candidates. The results of fits to the mass distributions are then shown for the $\mathrm{D}^0 \to \mathrm{K}^{-}\pi^{+}$ and $\overline{\mathrm{D}}{}^0 \to \mathrm{K}^{+}\pi^{-}$ datasets. These are found to contain $\sim$99.2 % signal (including secondary D^0), of which $\sim$95.8 % has the D^0 correctly reconstructed.

Finally, the results of fits to the proper-decay-time distributions and the effective lifetimes obtained are presented. These are performed on reduced datasets in which the fraction of secondary D^0 is suppressed, leaving 226,110 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{\mp}\pi^{\pm}$ and 30,481 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{+}\mathrm{K}^{-}$ candidates. These are found to consist of $\sim$99.5 % prompt D^0. The resulting values of y_{CP} and A_{Γ}, and their statistical uncertainties, are then shown.

Chapter 6 presents stability verification studies for these measurements, and evaluates their systematic uncertainties. Various cross-checks on the measured values of y_{CP} and A_{Γ} performed by dividing the data into subsets, showing the results to be stable. The fit method is also proven to give no significant bias via studies on Monte Carlo simulated toy data. Many sources of systematic uncertainties are evaluated, the dominant contributions to which come from combinatorial backgrounds and the parametrisation of the background from secondary D^0. The final results are found to be

$$\begin{aligned} A_{\Gamma} = &\quad (-5.9 \pm 5.9\,(\mathrm{stat.}) \pm 2.1\,(\mathrm{syst.})) \times 10^{-3}, \\ y_{CP} = &\quad (5.5 \pm 6.3\,(\mathrm{stat.}) \pm 4.1\,(\mathrm{syst.})) \times 10^{-3}. \end{aligned}$$

y_{CP} is consistent with the world average of y and with zero, and A_{Γ} is consistent with zero. Thus, these results show no evidence for CP-violation or mixing in the D^0 system.

Contents

List of Tables

List of Figures

Chapter 1

The Mathematical Theory of Particle Physics

From the very first description of the movement of the planets by the Babylonians to particle colliders like the Large Hadron Collider (LHC), the purpose of any physics experiment has always been to observe natural phenomena and attempt to understand them. This understanding now takes the form of a mathematical theory to describe the underlying physical mechanisms, which can then be used to predict future behaviour. Historically, observation has often preceded understanding - an experiment was performed and then the mathematical theory developed to explain its results. For instance, Galileo studied falling objects before Newton formalised their behaviour in the theory of gravity. However, in recent decades this situation has been reversed in particle physics. Particle physics examines and describes the nature and interactions of the most fundamental building blocks of the universe: elementary particles. The 1960s and 1970s saw the first high energy particle accelerator experiments performed and gave light to a plethora of newly observed particles and phenomena. Consequently, through the large collaborative effort of many theoretical physicists, the 1970s gave birth to the over-arching mathematical theory that describes these observations: what has come to be known as the Standard Model of particle physics (SM) [5].

The SM is one of the most successful physics theories in history. The results of almost every particle physics experiment performed to date have been consistent with the predictions of the SM. However, there are a small number of exceptions to this that call into question the completeness of the SM. The observation that neutrinos oscillate between types indicates that neutrinos have non-zero mass, which contradicts the SM. Indeed, the fact that there are three different types of neutrinos, as well as three types of quarks and charged leptons, is not explained by the SM. The SM also fails to explain indirect evidence for cosmological phenomena like dark matter and dark energy. Further, although the SM allows for some

difference in the behaviour of matter and anti-matter, it is insufficient to account for the dominance of matter in the universe. Thus, the SM is assumed to be an 'effective theory', which is mostly correct within the scope of current measurements, but might not explain physics at higher energies. Theoretical physicists have thus developed many possible extensions to the SM which agree with the SM at relatively low energies, but offer solutions to its known short-comings, and would also predict new physical phenomena at higher energies. Among these the most popular, and predictive, is a Supersymmetric theory (SUSY) [6]. In order for such a theory to solve the known issues in the SM its effects would have to be detectable at energies not much higher than those already covered by the SM.

The purpose of experiments like those of the LHC is thus to test the SM to its limits and attempt to discover new phenomena that it cannot explain. Any result contradicting the SM can be examined in the context of new physics theories to determine which of them actually describes particle physics. Further, such a result can allow these theories to make more precise predictions of as yet unobserved phenomena. It is for this reason that the LHC has pushed the limits in both energy and instantaneous luminosity. Higher energy allows the possible direct discovery of new particles outwith the scope of the SM, while high luminosity provides large quantities of data which can be used to test the predictions of the SM at much higher precision. The analyses presented in this thesis use data collected by the LHCb experiment at the LHC. LHCb is an experiment of the latter kind, aiming to reveal signs of new physics by making high precision measurements, and is described in detail in chapter 2.

The purpose of this chapter is to describe the SM and how it is used to make predictions of phenomena in particle physics, and thus motivate the measurements presented in chapters 4, 5 and 6. After an overview of the SM in section 1.1 a more focused description of the theory of the physical phenomena that are examined by LHCb is given in section 1.2. Finally, the theory behind the rate with which the D^0 charmed meson decays to two light mesons is discussed in section 1.3. The methodology and results of measuring this decay rate are presented in chapters 4 and 5 respectively, while the systematic uncertainties on these measurements are evaluated in chapter 6.

1.1 Overview of the Standard Model

1.1.1 Particle Content of the Standard Model

The most basic components of the universe are the elementary particles, which have been confirmed experimentally to have no substructure down to a scale of $\sim 10^{-19}\ \mathrm{m}$ [5]. There

Lepton	Mass [MeV/c^2]	Spin	Electric Charge	I_z	Colour Charge	e No.	μ No.	τ No.
Electron, e	0.511	$\frac{1}{2}$	-1	$-\frac{1}{2}$	0	$+1$	0	0
e Neutrino, ν_e	$< 2 \times 10^{-6}$	$\frac{1}{2}$	0	$+\frac{1}{2}$	0	$+1$	0	0
Muon, μ	105.65	$\frac{1}{2}$	-1	$-\frac{1}{2}$	0	0	+1	0
μ Neutrino, ν_μ	< 0.19	$\frac{1}{2}$	0	$+\frac{1}{2}$	0	0	+1	0
Tau, τ	$1,776.82 \pm 0.16$	$\frac{1}{2}$	-1	$-\frac{1}{2}$	0	0	0	+1
τ Neutrino, ν_τ	< 18.2	$\frac{1}{2}$	0	$+\frac{1}{2}$	0	0	0	+1

(a)

Quark	Mass [MeV/c^2]	Spin	Electric Charge	I_z	Colour Charge	I_3	C	S	T	B
Up, u	$2.3^{+0.7}_{-0.5}$	$\frac{1}{2}$	$+\frac{2}{3}$	$+\frac{1}{2}$	r, g or b	$+\frac{1}{2}$	0	0	0	0
Down, d	$4.8^{+0.7}_{-0.3}$	$\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$	r, g or b	$-\frac{1}{2}$	0	0	0	0
Charm, c	$1,275 \pm 25$	$\frac{1}{2}$	$+\frac{2}{3}$	$+\frac{1}{2}$	r, g or b	0	+1	0	0	0
Strange, s	95 ± 5	$\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$	r, g or b	0	0	-1	0	0
Top, t	$173,500 \pm 1,000$	$\frac{1}{2}$	$+\frac{2}{3}$	$+\frac{1}{2}$	r, g or b	0	0	0	+1	0
Bottom, b	$4,180 \pm 30$	$\frac{1}{2}$	$-\frac{1}{3}$	$-\frac{1}{2}$	r, g or b	0	0	0	0	-1

(b)

Boson	Mass [MeV/c^2]	Spin	Electric Charge	I_z	Colour Charge
Photon, γ	0	1	0	0	0
Z^0	$91,187.6 \pm 2.1$	1	0	0	0
$\mathrm{W}^\pm$	$80,385 \pm 15$	1	± 1	± 1	0
Gluon, g	0	1	0	0	8 combinations

(c)

Table 1.1: The properties of (a) the quarks, (b) the leptons and (c) the gauge bosons of the standard model. Here I_z is the z component of the weak isospin of the left-handed field. All right-handed fields have $I_Z = 0$. I_3 is the z component of the quark isospin, C is 'charm number', S is 'strangeness number', T is 'topness number', and B is 'bottomness number'. Gluons carry both colour and anti-colour charge, of which 8 different combinations that are not colour neutral are possible. Anti-particles have the same mass as particles with the sign of all quantum numbers reversed, except spin. The properties are obtained from [7].

are several different species of elementary particle, as shown in table 1.1. Each has a specific set of quantum numbers that determine how particles of that type interact with other particles. In addition, each type of particle has a corresponding anti-particle which has the same mass but opposite quantum numbers. One such quantum number is the intrinsic angular momentum, or 'spin'. The elementary particles can be broadly grouped into two categories depending on their spin: fermions, that have half-integer spin, and bosons, that have integer spin. This grouping is important due to the Spin Statistics Theorem [8], which states that fermion wavefunctions are anti-symmetric under the interchange of two identical particles, while boson wavefunctions are symmetric under such an interchange. A consequence of this is that fermions obey the Pauli exclusion principal, which states that no two identical fermions can occupy the same quantum state.

1.1.1.1 Standard Model Fermions

The group of fermions can again be grouped into two subsections: quarks and leptons. As mentioned previously, one of the natural phenomena that the SM cannot explain is that quarks and leptons have three 'generations'. The particles in each generation are identical except in their masses.

There are two leptons in each generation: one massive particle with electric charge -1, and one neutral particle with very little mass - a 'neutrino' (literally 'little neutral one'). In the first, lightest generation these are the electron (e^-) and the electron neutrino (ν_e). The second and third generations consist of the muon (μ^-) and tau (τ^-) and their corresponding neutrinos (ν_μ and ν_τ). The anti-particles to the charged leptons have charge +1 and are denoted by e^+, μ^+, and τ^+, and the anti-neutrinos by $\overline{\nu}_e$, $\overline{\nu}_\mu$, and $\overline{\nu}_\tau$. Each generation of lepton also has a 'flavour' quantum number. These are 'electron-number', 'muon-number' and 'tau-number' for the three generations respectively. The e^- and ν_e have +1 electron-number, while the e^+ and $\overline{\nu}_e$ have -1 electron-number, and all four have 0 muon- and tau-number. Likewise for the second and third generations. Lepton-number is conserved in all interactions, excluding neutrino oscillations.

The masses of each type of lepton are also shown in table 1.1. The neutrinos have such small masses that only upper limits on their masses have been measured. Neutrino masses are known to be non-zero due to the observation that they 'oscillate' between generations [9] - that is, they spontaneously change their type between generations, and thus violate conservation of lepton-number. More precisely, the states in which neutrinos propagate - their 'mass eigenstates' - are in fact superpositions of the states in which they interact. Thus, if a neutrino of one type is produced in an interaction the probability that it then interacts as a given type oscillates with time. The rate of this oscillation is dependent on the difference in

mass between the neutrino types, thus if neutrinos were massless no oscillation would occur. However, all calculations in the SM are performed under the assumption that neutrinos are massless. As neutrino masses are very small compared to those of the other elementary particles this is a very good approximation. The predictive power of the SM with respect to the physical phenomena discussed here is thus unaffected.

Similar to the leptons, there are two quarks in each generation: one 'up-type' with electric charge $+\frac{2}{3}$ and one 'down-type' with electric charge $-\frac{1}{3}$. These names come from the first, lightest generation, which consists of the up (u) and down (d) quarks. The second generation consists of the charm (c) and strange (s) quarks, and the third generation the top (t) and bottom (b) quarks (also called the truth and beauty quarks, by the more poetic physicists). A recent result from the ATLAS experiment at the LHC has confirmed that quarks have no sub-structure down to $\sim 6 \times 10^{-20}\,\mathrm{m}$ [10].

Where quarks differ from leptons is that they also carry 'colour charge'. Colour charge can take three values: red, blue and green. Its existence was first indicated experimentally by the discovery of the Ω^- [11], which is a bound state of three d quarks. Such a state would be forbidden by the Pauli exclusion principal if not for the existence of an additional quantum number that takes three values to distinguish the otherwise identical quarks. Anti-quarks have opposite electric charge, and opposite colour charge, which takes values anti-red, anti-blue or anti-green.

Another parallel to leptons is that quarks also have a flavour quantum number. This is isospin for the first generation, charm and strangeness for the second, and topness and bottomness (or beauty, if you prefer) for the third. Unlike the leptons in the SM, quark flavour is not strictly conserved in all interactions, and it is this aspect of the quark sector that makes it so interesting to study. The consequences of this feature are the motivation for the analyses presented in this thesis, and will be discussed further in sections 1.2 and 1.3.

1.1.1.2 Standard Model Bosons

There are four fundamental forces of nature: the strong, electromagnetic, weak, and gravitational forces. Each force has an associated set of bosons that mediate the transfer of momentum between particles [12]. These have been discovered experimentally for the strong, electromagnetic and weak forces, but not for gravity. Gravity is by far the weakest of the fundamental forces. It can safely be neglected when considering high energy interactions of elementary particles, and so will not be discussed any further.

The boson associated with the electromagnetic force is the photon (γ), which couples to electric charge. Consequently all electrically charged particles, *i.e.* all fermions except neutrinos, experience the electromagnetic force. The photon is massless and neutral, which

means the photon cannot decay to lighter particles, or couple to itself, giving electromagnetism infinite range.

The weak force is mediated by the charged $\mathrm{W}^{\pm}$ and neutral Z^0 bosons, which couple to all fermions (at least all 'left-handed' fermions, as will be discussed in section 1.1.3.2). Their masses are relatively large: $M_W \simeq 80.4\,\mathrm{GeV}/c^2$ and $M_Z \simeq 91.2\ \mathrm{GeV}/c^2$. Consequently the range of the weak force is very small, $\mathcal{O}(10^{-18}\ \mathrm{m})$. As will be shown in section 1.1.3.2, the weak and electromagnetic forces are in fact manifestations of the underlying electroweak force, and the $\mathrm{W}^{\pm}$, Z^0 and γ the physical manifestations of its four gauge bosons.

The strong force is mediated by gluons (g), which couple to colour charge. Thus quarks experience the strong force while leptons do not. Gluons are massless, like the photon, but carry colour charge themselves, which means gluons can interact with other gluons, unlike photons. Also, as its name suggests, the strong force is by far the strongest of the fundamental forces. This means gluons can only travel very short distances before interacting, meaning the range of the strong force is only $\mathcal{O}(10^{-15}\ \mathrm{m})$ - roughly the radius of a nucleon.

This gives rise to the phenomenon of 'colour confinement' - only colour neutral states are stable, and observable. Consequently, no isolated quark has ever been observed. Instead, quarks exist in colour neutral bound states: either as a meson, which consists of a quark and an anti-quark with equal and opposite colour; or as a baryon, which consists of three quarks (or anti-quarks) each with different colour, causing it to be colour neutral. The quark and anti-quark in mesons can annihilate each other, meaning even the lightest mesons eventually decay. However, this is not the case for baryons. The lightest, and thus stable, baryon is the proton, which consists of uud valence quarks; second lightest is the neutron, which consists of udd valence quarks. Together these form atomic nuclei, and with the electron - the lightest charged lepton - make up the atoms that comprise all the stable, visible matter of the universe.

In their simplest form, the symmetries of the SM dictate that interactions should be identical regardless of which generation of fermion is involved. However, this is known experimentally not to be the case, due to the differing masses of the fermions in each generation, and of the $\mathrm{W}^{\pm}$ and Z^0 bosons. Thus the Higgs boson was proposed as an addition to the perfectly symmetric SM [13]. The Higgs boson breaks the symmetry of the SM, and gives the SM particles their different masses. An excess of signal which is consistent with being the Higgs boson has recently been observed by the LHC experiments ATLAS and CMS at a mass of $\sim$125 GeV/c^2 with a significance of 5σ [14, 15].

1.1.2 The Standard Model as a Gauge Theory

The bosons of the SM, and the exchanges of quantum numbers that they perform, can be represented as gauge groups. These describe the underlying symmetries of the SM, what

transitions between states are allowed or forbidden and thus how the elementary particles interact. The gauge groups specific to the SM are chosen to match experimental observation, and could be modified to account for any new, non-SM particles that may be discovered at the LHC.

An $SU(n)$ gauge group has $n^2 - 1$ generators, the physical interpretation of which is a fundamental force with $n^2 - 1$ gauge bosons [12]. The photon has no charge itself, and simply transfers momentum between charged particles. It is thus natural to assume that electromagnetism can be described by a $U(1)$ group. The weak force, on the other hand, has three bosons - the $\mathrm{W}^{\pm}$ and Z^0- and allows transitions between states with a charge difference of 1. Thus, it is likely to correspond to a $SU(2)$ group. However, it was found by Glashow, Salam and Weinberg that in fact the electromagnetic and weak forces can be described as the physical manifestations of the underlying 'electroweak' force, which has gauge group $SU(2) \times U(1)_Y$ [16, 17, 18]. The gauge bosons of this group are not equal to the γ, $\mathrm{W}^{\pm}$, and Z^0, but are instead the $\mathrm{W}^{1,2,3}$ for the $SU(2)$ group, and the B for the $U(1)_Y$ group. Their respective coupling constants are g and g'. It will be shown in the following section how these give rise to the bosons observed in nature.

The gluons that mediate the strong force allow transitions between states with three different colour charges. Thus, interactions via the strong force can be described by an $SU(3)$ group. Consequently there are 8 types of gluon, G^a for $a \in \{1, ..., 8\}$, that carry both colour and anti-colour charge, and have coupling constant g_s.

The whole SM is thus described by a gauge group $SU(3) \times SU(2) \times U(1)_Y$. In order to describe the interactions of the fundamental particles one must use Quantum Field Theory, which is described in the following section.

1.1.3 Interactions of Elementary Particles

The most general description of any dynamical system is the Lagrangian $\mathcal{L}$. The exact form of $\mathcal{L}$ depends on the spin of the particle that it describes. Applying the 'Principle of Least Action' [19] to the Lagrangian for a field ψ results in the Euler-Lagrange equation

$$\frac{\partial \mathcal{L}}{\partial \psi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial \left(\partial_\mu \psi \right)} \right) = 0, \tag{1.1}$$

where ∂_μ is the covariant derivative, the index μ running over the 4 space-time coordinates, from which one obtains the equations of motion for the system. The Lagrangian for a system of many different particles undergoing many different interactions is simply the sum of the Lagrangians for the individual particles and interactions. The SM is thus best described by its Lagrangian.

Quantum Field Theory, as its name suggests, describes all particles as mathematical fields. For a scalar (spin-0) field ϕ, such as the Higgs boson, the Lagrangian is [12]

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu - \frac{1}{2}m^2\phi^2 - \frac{\lambda}{4!}\phi^4, \tag{1.2}$$

where m is the mass of the field, and the final term represents an interaction with coupling λ. This is the simplest interaction term that can be added while keeping the theory renormalisable - a necessary requirement for any physical theory. Applying the Euler-Lagrange equation to this Lagrangian, in the absence of any interaction term, yields the Klein-Gordon equation.

The spin-$\frac{1}{2}$ fermions are described as Dirac spinor fields, ψ, as they must also carry information on the direction of the spin. Their Lagrangian is given by

$$\mathcal{L} = \overline{\psi}(i\gamma^\mu\partial_\mu - m)\psi, \tag{1.3}$$

where γ^μ are the Dirac γ-matrices, and $\overline{\psi} \equiv \psi^\dagger\gamma^0$. Applying the Euler-Lagrange equation here yields the Dirac equation.

Finally, spin-1 bosons are described by a vector field, A, for which the Lagrangian is

$$\mathcal{L} = -\frac{1}{4}F_{\mu\nu}F^{\mu\nu}, \tag{1.4}$$

where the field strength $F_{\mu\nu} \equiv \partial_\mu A_\nu - \partial_\nu A_\mu$, both for abelian fields, like the photon, and non-abelian fields, like gluons. If A represents the photon field, applying the Euler-Lagrange equation results in Maxwell's equations of electromagnetism.

The total SM Lagrangian comprises the components for each type of particle

$$\mathcal{L} = \mathcal{L}_{boson\,kinetic} + \mathcal{L}_{fermion\,kinetic} + \mathcal{L}_{fermion\,masses} + \mathcal{L}_{Higgs}. \tag{1.5}$$

The first term describes the kinematics of the bosons, and the second term likewise for the fermions; the third term describes the coupling of the Higgs field to the fermions, thus generating their masses, and the final term likewise for the bosons.

1.1.3.1 Boson Kinetic Term

Following the form of equation 1.4 and adding a term for each of the SM bosons one obtains

$$\mathcal{L}_{boson\,kinetic} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}F^a_{\mu\nu}F^{a\,\mu\nu} - \frac{1}{4}F^A_{\mu\nu}F^{A\,\mu\nu} + \mathcal{L}_{gauge\,fixing} + \mathcal{L}_{ghosts}, \tag{1.6}$$

where $B_{\mu\nu}$ is the field strength for the B field, corresponding to $U(1)_Y$, $F^a_{\mu\nu}$ are those for the $SU(2)$ W^a bosons, with $a \in \{1,2,3\}$, and $F^A_{\mu\nu}$ those for the $SU(3)$ gluons, with $a \in \{1,...,8\}$. The second last term allows a specific gauge to be chosen, which results in extra 'Fadeev-Popov ghost' bosons. These particles are allowed to enter in loop processes, but are cancelled out by loops of gauge bosons, and thus do not contribute to any observables.

1.1.3.2 Fermion Kinetic Term

The kinetic term of the Lagrangian for fermions consists of both the terms for all fermions interacting via the electroweak force, and for the quarks interacting via the strong force.

Considering the weak force first, it is known to maximally violate parity, and only couples to 'left-handed' fields. The 'handedness' of a vector field specifies whether its spin points in the same direction as its momentum (right-handed), or the opposite direction (left-handed). For a vector field ψ the projection operators $P_{L/R} \equiv \frac{1\mp\gamma^5}{2}$ separate out the left- (right-) handed components of the field. Thus, any field can be written as a superposition of left- and right-handed components:

$$\psi = P_L\psi + P_R\psi \equiv \psi_L + \psi_R. \tag{1.7}$$

The interactions of these two components can thus be treated separately.

Considering only the first generation of fermions, the left-handed fields form $SU(2)$ doublets

$$\mathrm{q}_L \equiv \begin{pmatrix} \mathrm{u}_L \\ \mathrm{d}_L \end{pmatrix} \text{ and } \ell_L \equiv \begin{pmatrix} \nu_L \\ \mathrm{e}_L \end{pmatrix}, \tag{1.8}$$

and the right-handed fields form $SU(2)$ singlets

$$\mathrm{u}_R,\ \mathrm{d}_R,\ \nu_R \text{ and } \mathrm{e}_R. \tag{1.9}$$

As the neutrino has no electric charge, and right-handed fields cannot interact via the weak force, the right-handed neutrino cannot interact at all. Hence, it is ignored in the following discussion. In fact, under the assumption that neutrinos are massless they have definite chirality: either left- or right-handed. As the right-handed neutrino does not interact none can ever be produced, leading to the conclusion that right-handed neutrinos do not exist (and conversely, neither do left-handed anti-neutrinos).

Each of these doublets and singlets thus contributes a term of the form of equation 1.3 to the Lagrangian (excluding the mass term which will be discussed in the following section):

$$\begin{aligned} \mathcal{L}_{fermion\,kinetic} = & i\overline{\ell_L}^T\gamma^\mu \mathbf{D}_\mu^{\ell_L}\ell_L + i\overline{\mathrm{e}_R}\gamma^\mu D_\mu^{\mathrm{e}_R}\mathrm{e}_R \\ & + i\overline{\mathrm{q}_L}^T\gamma^\mu \mathbf{D}_\mu^{\mathrm{q}_L}\mathrm{q}_L + i\overline{\mathrm{d}_R}\gamma^\mu D_\mu^{\mathrm{d}_R}\mathrm{d}_R + i\overline{\mathrm{u}_R}\gamma^\mu D_\mu^{\mathrm{u}_R}\mathrm{u}_R \\ & + \mathcal{L}_{quark\,strong\,interaction}, \end{aligned} \tag{1.10}$$

where the couplings to the relevant gauge bosons are included in the covariant derivatives. The term for the strong force, $\mathcal{L}_{quark\,strong\,interaction}$, is discussed later in this section. The right-handed singlets couple only to the $U(1)_Y$ boson, thus

$$D_\mu^{f_R} \equiv \partial_\mu + ig'Y(f_R)\mathrm{B}_\mu, \tag{1.11}$$

where the 'weak hypercharge' Y is given by

$$Y(f) \equiv 2(Q(f) - I_Z(f)), \tag{1.12}$$

with $Q(f)$ the electric charge of the fermion f, and $I_Z(f)$ its weak isospin, with $I_Z(\mathrm{u}_L) = I_Z(\nu_L) = +\frac{1}{2}$, $I_Z(\mathrm{d}_L) = I_Z(\mathrm{e}_L) = -\frac{1}{2}$ and $I_Z(f_R) = 0$.

The left handed leptons couple to the $SU(2) \times U(1)$ bosons and so

$$\mathbf{D}_\mu^{\ell_L} \equiv \partial_\mu \mathbf{I} + ig\mathbf{T}^a \mathrm{W}_\mu^a + ig'\mathbf{Y}(\ell_L)\mathrm{B}_\mu, \tag{1.13}$$

where $\mathbf{T}^a$ are the $SU(2)$ generators, one representation of which is the set of Pauli spin matrices. Expanding the first term of equation 1.10 using equation 1.13 the interaction terms give

$$-\frac{g}{2}\begin{pmatrix}\overline{\nu}_L & \overline{\mathrm{e}}_L\end{pmatrix}\gamma^\mu\left[\begin{pmatrix}\mathrm{W}_\mu^3 & \mathrm{W}_\mu^1 + i\mathrm{W}_\mu^2 \\ \mathrm{W}_\mu^1 - i\mathrm{W}_\mu^2 & -\mathrm{W}_\mu^3\end{pmatrix} + \frac{g'}{g}\mathrm{B}_\mu\right]\begin{pmatrix}\nu_L \\ \mathrm{e}_L\end{pmatrix}. \tag{1.14}$$

Requiring that there be two physical charged bosons ($\mathrm{W}^\pm$) with no mixing (orthogonal states) one can read off from the off diagonal terms to get

$$\mathrm{W}^\pm = \frac{1}{\sqrt{2}}(\mathrm{W}^1 \mp i\mathrm{W}^2). \tag{1.15}$$

For the neutral bosons, introducing the Weinberg angle θ_W such that $\tan\theta_\mathrm{W} = \frac{g'}{g}$, one can write

$$\begin{pmatrix}\mathrm{Z}^0 \\ A\end{pmatrix} \equiv \begin{pmatrix}\cos\theta_\mathrm{W} & -\sin\theta_\mathrm{W} \\ \sin\theta_\mathrm{W} & \cos\theta_\mathrm{W}\end{pmatrix}\begin{pmatrix}\mathrm{W}^3 \\ \mathrm{B}\end{pmatrix}. \tag{1.16}$$

Thus, $SU(2)$ and $U(1)_Y$ are unified in the physical manifestations of their bosons. Rewriting equation 1.14 in terms of the physical bosons gives

$$-\frac{g}{2}\begin{pmatrix}\overline{\nu}_L & \overline{\mathrm{e}}_L\end{pmatrix}\gamma^\mu\begin{pmatrix}\frac{\mathrm{Z}_\mu^0}{\cos\theta_\mathrm{W}} & \sqrt{2}\mathrm{W}_\mu^- \\ \sqrt{2}\mathrm{W}_\mu^+ & \cos\theta_\mathrm{W}\mathrm{Z}_\mu^0 - 2\sin\theta_\mathrm{W}A_\mu\end{pmatrix}\begin{pmatrix}\nu_L \\ \mathrm{e}_L\end{pmatrix}. \tag{1.17}$$

This shows that a ν-ν vertex can only couple to the Z^0, while an e-e vertex can couple to either the Z^0 or γ. Thus, the γ couples only to electrically charged particles.

In addition to the electroweak interactions, the quarks also undergo strong interactions. These are blind to the handedness of the quark fields, and instead couple to their colour charge. Thus one must express the quark fields in terms of the three fields of different colour charge, which form $SU(3)$ triplets:

$$\mathrm{u}_\mathbf{c} = \begin{pmatrix}\mathrm{u}_r \\ \mathrm{u}_g \\ \mathrm{u}_b\end{pmatrix} \text{ and } \mathrm{d}_\mathbf{c} = \begin{pmatrix}\mathrm{d}_r \\ \mathrm{d}_g \\ \mathrm{d}_b\end{pmatrix}, \tag{1.18}$$

where the subscripts r, g and b represent the colour charges, red, green and blue. Writing the Lagrangian interaction term in exactly the same way as the electroweak interactions one thus obtains

$$\mathcal{L}_{quark\,strong\,interaction} = -g_s \overline{\mathrm{u}}_{\mathbf{c}}{}^T \gamma^\mu G_\mu^A \mathbf{T_s^A} \mathrm{u}_{\mathbf{c}} - g_s \overline{\mathrm{d}}_{\mathbf{c}}{}^T \gamma^\mu G_\mu^A \mathbf{T_s^A} \mathrm{d}_{\mathbf{c}}, \tag{1.19}$$

where G^A are the gluon fields, and $\mathbf{T_s^A}$ the generators of $SU(3)$, with $A \in \{1, ..., 8\}$. Summing over the gluon fields the generators $\mathbf{T_s^A}$ describe which quark colour states can couple to which gluons. However, as the gluons are massless, their different physical manifestations are indistinguishable. Consequently, no such transformation of the underlying states to the physical states, as was done for the electroweak bosons, can or need be performed.

1.1.3.3 Fermion Masses

An explicit mass term for fermions of the form $m\overline{\psi}\psi = m(\overline{\psi}_R\psi_L + \overline{\psi}_L\psi_R)$ breaks the $SU(2)$ invariance of the Lagrangian, as the left- and right-handed fields form different multiplets under $SU(2)$, and so cannot be mixed. This can be solved by introducing a scalar field Φ that forms an $SU(2)$ doublet, and has a non-zero vacuum expectation value (VEV) - the 'Higgs field' [13]. The interactions of the fermions with this field take the form of a mass term

$$\mathcal{L}_{Yukawa} = -Y_f \overline{f}_L^T \Phi f_R + \mathrm{h.c.}, \tag{1.20}$$

where Y_f is the 'Yukawa coupling', and 'h.c.' is the hermitian conjugate of the first term.

The Higgs potential has the form

$$V(\Phi) = -\mu^2 \Phi^* \Phi + \lambda |\Phi^* \Phi|^2, \tag{1.21}$$

with $\mu^2, \lambda > 0$. The minimum potential occurs at $\Phi^*\Phi = \frac{1}{2}\mu^2/\lambda \equiv \frac{1}{2}v^2$, where v is the VEV. In the unitary gauge only one component of the Higgs doublet obtains the VEV, which can be written as

$$\langle \Phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix}. \tag{1.22}$$

Thus one can simply write

$$\Phi = \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{1.23}$$

and substituting this into equation 1.20 the leptonic part gives

$$\mathcal{L}_{Yukawa} = \frac{-Y_f}{\sqrt{2}} \begin{pmatrix} \overline{\nu}_L & \overline{\mathrm{e}}_L \end{pmatrix} \begin{pmatrix} 0 \\ v + H \end{pmatrix} \mathrm{e}_R + \mathrm{h.c.}. \tag{1.24}$$

Thus, the mass term for the electron is

$$\begin{aligned} -\frac{Y_e v}{\sqrt{2}}(\bar{\mathrm{e}}_L \mathrm{e}_R + \bar{\mathrm{e}}_R \mathrm{e}_L) &= \frac{Y_e v}{\sqrt{2}} \bar{\mathrm{e}}\mathrm{e} \\ &\equiv m_{\mathrm{e}} \bar{\mathrm{e}}\mathrm{e}, \end{aligned} \tag{1.25}$$

so the electron obtains a mass proportional to the Higgs VEV, while retaining the $SU(2)$ invariance of the Lagrangian. The terms involving H represent the interactions between the Higgs boson and the electron, which are also proportional to m_{e}. Repeating this for the d quark yields its mass term.

As the upper component of Φ is zero this doesn't yield a mass term for the u quark. However, this is achieved by exploiting $SU(2)$ symmetry, and including a term like equation 1.20 but with Φ replaced by

$$\begin{aligned} \Phi^c &= \epsilon_{ij} \Phi^{j*} \\ &= \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v + H^* \end{pmatrix} \\ &= \begin{pmatrix} v + H^* \\ 0 \end{pmatrix}, \end{aligned} \tag{1.26}$$

which naturally yields a mass term for the u

$$m_{\mathrm{u}} = \frac{Y_{\mathrm{u}}}{\sqrt{2}} v. \tag{1.27}$$

1.1.3.4 Boson Masses

As the Higgs boson is scalar its contribution to the Lagrangian takes the form of equation 1.2. Incorporating its interactions with the $SU(2)$ bosons into the covariant derivative yields

$$\mathbf{D}_\mu \Phi = \frac{1}{\sqrt{2}} \left(\partial_\mu + i\frac{g}{2} \begin{pmatrix} \mathrm{W}^3_\mu & \sqrt{2}\mathrm{W}^-_\mu \\ \sqrt{2}\mathrm{W}^+_\mu & -\mathrm{W}^3_\mu \end{pmatrix} + i\frac{g'}{2}\mathrm{B}_\mu \right) \begin{pmatrix} 0 \\ v + H \end{pmatrix}, \tag{1.28}$$

which gives

$$|\mathbf{D}_\mu|^2 = \frac{1}{2}(\partial_\mu H)^2 + \frac{g^2 v^2}{4} \mathrm{W}^{+\mu}\mathrm{W}^-_\mu + \frac{v^2}{8}\left(g\mathrm{W}^3_\mu - g'\mathrm{B}_\mu\right)^2 + \text{interaction terms}, \tag{1.29}$$

where 'interaction terms' refers to those terms concerning the coupling of the Higgs to other bosons. Substituting in equation 1.16 gives the mass terms of the physical $\mathrm{W}^\pm$, Z^0 and γ

$$|\mathbf{D}_\mu|^2 = \frac{1}{2}(\partial_\mu H)^2 + \frac{g^2 v^2}{4} \mathrm{W}^{+\mu}\mathrm{W}^-_\mu + \frac{v^2 g^2}{8\cos^2\theta_{\mathrm{W}}} \mathrm{Z}^0_\mu \mathrm{Z}^{0\mu} + 0 A_\mu A^\mu. \tag{1.30}$$

Here we see that the $\mathrm{W}^\pm$ and Z^0 have acquired masses

$$m_{\mathrm{W}^\pm} = \frac{1}{2} g v, \text{ and } m_{\mathrm{Z}^0} = \frac{1}{2}\frac{g v}{\cos\theta_{\mathrm{W}}}, \tag{1.31}$$

which are related by $m_{\mathrm{W}^{\pm}} = \cos\theta_{\mathrm{W}} m_{\mathrm{Z}^0}$. The γ has remained massless, again showing that the unification of $SU(2)$ and $U(1)$ neatly describes the observable bosons. Further, as Φ is a singlet under $SU(3)$ it does not interact with gluons, which consequently remain massless as well.

1.1.4 Additional Generations

The previous section described the interactions of two quarks, u and d, and two leptons, e, and ν_e, via the $SU(3) \times SU(2) \times U(1)_L$ gauge bosons. However, as was described in section 1.1.1, there are in fact two further generations of fermions that are identical to this first generation, but contain particles with larger mass.

Due to their identical quantum numbers the couplings of these additional particles to the gauge bosons are identical to their first generation counterparts, except in the Yukawa couplings that give rise to their mass. The couplings for each generation can be chosen to reproduce the measured masses of the fermions. However, the Yukawa couplings can be expressed as matrices in flavour space, so the Lagrangian, equation 1.20, becomes

$$\mathcal{L}_{Yukawa} = -[Y_{\mathrm{d}}]_{ij} \overline{\mathrm{d}}^{T}_{Li} \Phi \mathrm{d}_{Rj} - [Y_{\mathrm{u}}]_{ij} \overline{\mathrm{u}}^{T}_{Li} \Phi^{\mathrm{c}} \mathrm{u}_{Rj} + \mathrm{h.c.}, \tag{1.32}$$

where the indices i and j are summed over the generations. Any non-zero off diagonal elements of the Yukawa matrices gives rise to mixing between the generations.

As the interactions of the particles one generation are the same as those in any other generation the fermion kinetic Lagrangian term is invariant under unitary rotations between generations. Thus, one can rotate the flavour eigenstates to give the mass eigenstates, which correspond to the diagonalised Yukawa matrices. Consequently, the states with definite mass, in which the quarks propagate, are generally superpositions of those with definite flavour, in which the quarks interact. Considering only the first two generations of quark, the most general 2×2 unitary matrix can be written as

$$\mathbf{V_C} = \begin{pmatrix} \cos\theta_C & \sin\theta_C \\ -\sin\theta_C & \cos\theta_C \end{pmatrix}, \tag{1.33}$$

where θ_C determines the level of mixing between the two generations. Three complex phases can also enter into such a matrix, but these can be removed by phase transformations of the u, d and s fields, leaving only one free parameter. Then the mass eigenstates of the d and s can be expressed as

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \end{pmatrix} = \mathbf{V_C} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \end{pmatrix}. \tag{1.34}$$

This mixing transformation was first proposed by Cabibbo [20], at a time when only the u, d and s quarks had been discovered experimentally, and resulted in the prediction of the existence of the c quark. θ_C has been measured to be $(13.04 \pm 0.05)^\circ$ [7], thus the mixing between d and s is large, but not maximal.

Consequently, the coupling of the d and s to the W^- can be written as

$$-\frac{g}{2\sqrt{2}}\begin{pmatrix}\overline{\mathrm{u}} & \overline{c}\end{pmatrix}\gamma^\mu(1-\gamma^5)\mathbf{V_C}\begin{pmatrix}\mathrm{d}\\ \mathrm{s}\end{pmatrix}\mathrm{W}^-_\mu + \mathrm{h.c.}, \tag{1.35}$$

which shows that a u quark can undergo an interaction in which it is transformed to a s quark, and a c quark can be transformed into a d. However, the amplitude of these interactions is suppressed by a factor of $\sin\theta_\mathrm{W}$, and so are 'Cabibbo suppressed'. Those proportional to $\cos\theta_\mathrm{W}$ are 'Cabibbo favoured'.

In neutral interactions, via the exchange of a Z^0 or γ, the coupling is

$$-\frac{g}{2\sqrt{2}}\begin{pmatrix}\overline{\mathrm{d}} & \overline{\mathrm{s}}\end{pmatrix}\mathbf{V_C}^T\gamma^\mu(1-\gamma^5)\mathbf{V_C}\begin{pmatrix}\mathrm{d}\\ \mathrm{s}\end{pmatrix}\mathrm{Z}^0_\mu + \mathrm{h.c.}, \tag{1.36}$$

and as $\mathbf{V_C}^T = \mathbf{V_C}^{-1}$ the mixing terms cancel out. Thus there are no flavour changing neutral currents in the SM. This phenomenon was first predicted by Glashow, Iliopoulos and Maiani [21] and is thus known as the GIM mechanism.

When the third generation is included the mixing matrix becomes

$$\begin{aligned}\mathbf{V_{CKM}} &= \begin{pmatrix}\cos\theta_{12} & \sin\theta_{12} & 0\\ -\sin\theta_{12} & \cos\theta_{12} & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\theta_{23} & \sin\theta_{23}\\ 0 & -\sin\theta_{23} & \cos\theta_{23}\end{pmatrix}\begin{pmatrix}\cos\theta_{13} & 0 & \sin\theta_{13}e^{-i\delta}\\ 0 & 1 & 0\\ -\sin\theta_{13}e^{i\delta} & 0 & \cos\theta_{13}\end{pmatrix}\\ &= \begin{pmatrix}c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta}\\ -s_{12}c_{23}-c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23}-s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13}\\ s_{12}s_{23}-c_{12}c_{23}s_{13}e^{i\delta} & -c_{12}c_{23}-s_{12}s_{23}s_{13}e^{i\delta} & c_{23}c_{13}\end{pmatrix},\end{aligned} \tag{1.37}$$

where $\theta_{12} = \theta_C$, and the shorthand $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$ is used. Similarly to the Cabibbo mixing matrix the three angles θ_{ij} determine the level of mixing between the three pairs of generations. However, in the 3×3 case 6 complex phases can enter, only 5 of which can be removed by phase transformations of the quark fields - thus one complex phase remains in the mixing matrix. This allows for the phenomenon of CP-violation, which is discussed in section 1.2. Observation of CP-violation thus caused Kobayashi and Maskawa to propose the existence of the third generation before its experimental discovery [22]. Thus, the Cabibbo mixing matrix was extended into what has come to be known as the Cabibbo-Kobayashi-Maskawa, or CKM, matrix.

The mass eigenstates are given by

$$\begin{pmatrix} \mathrm{d}' \\ \mathrm{s}' \\ \mathrm{b}' \end{pmatrix} = \mathbf{V_{CKM}} \begin{pmatrix} \mathrm{d} \\ \mathrm{s} \\ \mathrm{b} \end{pmatrix}, \tag{1.38}$$

and consequently the matrix can also be written

$$\mathbf{V_{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix}, \tag{1.39}$$

with $|V_{ij}|^2$ giving the probability of a transition $i \to j$. The best current measurements of the mixing angles and CP-violating phase, in addition to θ_C, are [7]

$$\theta_{13} = (0.201 \pm 0.011)^{\circ},\ \theta_{23} = (2.38 \pm 0.06)^{\circ},\ \text{and}\ \delta = 1.20 \pm 0.08. \tag{1.40}$$

The same form of mixing matrix exists for leptons, and is the source of the neutrino oscillations discussed in section 1.1.1.1. The ramifications of the complex phase in the CKM matrix, and its allowance of CP-violation in transitions of quarks between generations, is the topic of flavour physics, which is discussed in the next section.

1.2 Flavour Physics and CP-Violation

1.2.1 Discrete Symmetries

The symmetry of the SM Lagrangian under the continuous transformations represented by the gauge bosons gives rise to the conservation of the charges to which they couple, *i.e.* electric charge, weak isospin, and colour. It is also symmetric under translations and rotations in space and time, which gives rise to momentum and energy conservation. Such continuous symmetries are key to describing the particles and forces of nature.

There also exist discrete transformations that give rise to important symmetries in nature. Three such transformations are charge conjugation (C), which reverses the signs of internal quantum numbers, transforming particle to anti-particle; parity (P) which reverses the spatial coordinates, and consequently switches the chirality of a particle; and time reversal (T), which transforms $t \to -t$ and so reverses the linear and angular momentum of a particle. Any theory that respects Lorentz invariance must also be invariant under the combination of these three transformations, CPT, but need not necessarily be invariant under each individual transformation. One consequence of CPT invariance is that the masses of a particle and its corresponding anti-particle must be identical.

It is theorised that matter and anti-matter were created in equal quantities in the early universe, thus one of the major outstanding puzzles of the universe is the dominance of matter over anti-matter that exists today. This implies that there is some violation of invariance under the exchange of particle for anti-particle. The weak interaction is known to maximally violate C and P, as it only couples to left-handed fermions and right-handed anti-fermions, but this does not explain matter anti-matter asymmetry. This leaves the combination CP, which, under CPT invariance, corresponds to T. As has been mentioned previously the non-zero complex phase in the CKM matrix allows CP-violation (or more precisely, violation of CP invariance) in charged weak interactions. That said, the level of CP-violation permitted by the CKM mechanism is many orders of magnitude too small to account for all of the matter-anti-matter asymmetry in the universe today. However, study of CP-violation may reveal new physics mechanisms to explain this.

1.2.2 CP-Violation in the CKM Matrix

An alternative representation of the CKM matrix is the Wolfenstein parametrisation [23] which defines [7]

$$\begin{aligned} \lambda &\equiv \sin\theta_{12} &&\simeq 0.23 \\ A &\equiv \frac{\sin\theta_{23}}{\sin^2\theta_{12}} &&\simeq 0.81, \text{ and} \\ \rho - i\eta &\equiv \frac{\sin\theta_{13}e^{-i\delta}}{\sin\theta_{12}\sin\theta_{23}} &&\simeq 0.14 - 0.35i, \end{aligned}$$

and expands up to $\mathcal{O}(\lambda^3)$ to give

$$\mathbf{V_{CKM}} = \begin{pmatrix} 1-\frac{1}{2}\lambda^2 & \lambda & A\lambda^3(\rho - i\eta) \\ -\lambda & 1-\frac{1}{2}\lambda^2 & A\lambda^2 \\ A\lambda^3(1-\rho-i\eta) & -A\lambda^2 & 1 \end{pmatrix} + \mathcal{O}(\lambda^4). \tag{1.41}$$

This shows the order of each of the elements more clearly: the diagonal elements are all close to 1, while the complex phase only enters at this order in transitions between the 1^{st} and 3^{rd} generations. Thus, D mesons, which consist of a c quark bound with a s, u or d, are expected to exhibit very little CP-violation. It is thus of great consequence for the SM, and new physics theories, that LHCb has recently observed evidence for direct CP-violation in decays of the D^0 meson at $\mathcal{O}(10^{-3})$ [1].

The unitarity of the CKM matrix requires that $\mathbf{V_{CKM}}\mathbf{V_{CKM}}^\dagger = \mathbf{I}$. Firstly this requires the mass eigenstates to be normalised

$$\Sigma_{j=\mathrm{d,s,b}} |V_{ij}|^2 = 1, \tag{1.42}$$

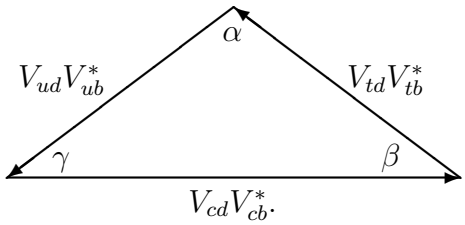


Figure 1.1: One of six triangles in the complex plane made by the unitarity constraint on the CKM matrix. This triangle, and its associated B decays, is most commonly studied as its angles are all of roughly the same size, and so is associated with the largest levels of CP-violation.

for $i \in \{\mathrm{u}, c, \mathrm{t}\}$. There are also six other constraints from the requirement that the off-diagonal elements of the product are zero. For example, one must have that

$$V_{ud}V_{ub}^* + V_{cd}V_{cb}^* + V_{td}V_{tb}^* = 0. \tag{1.43}$$

As each CKM element is a complex number each term of this equation forms one side of a triangle in the complex plane, with one corner at the origin, as shown in figure 1.1. The angles in this triangle are given by

$$\alpha = -\arg\left(\frac{V_{td}V_{tb}^*}{V_{ud}V_{ub}^*}\right), \tag{1.44}$$

$$\beta = -\arg\left(\frac{V_{td}V_{tb}^*}{V_{cd}V_{cb}^*}\right), \tag{1.45}$$

$$\gamma = -\arg\left(\frac{V_{ud}V_{ub}^*}{V_{cd}V_{cb}^*}\right), \tag{1.46}$$

each of which dictates the level of CP-violation in different transitions between quark generations.

The amplitudes A of transitions of a particle P or its anti-particle $\overline{\mathrm{P}}$, to a final state f or its charge conjugate $\overline{f}$ are given by

$$A_f = \langle f|\mathcal{H}|\mathrm{P}\rangle \qquad \bar{A}_{\bar{f}} = \langle \bar{f}|\mathcal{H}|\bar{\mathrm{P}}\rangle \tag{1.47}$$

$$A_{\bar{f}} = \langle \bar{f}|\mathcal{H}|\mathrm{P}\rangle \qquad \bar{A}_f = \langle f|\mathcal{H}|\bar{\mathrm{P}}\rangle, \tag{1.48}$$

where $\mathcal{H}$ is the Hamiltonian. Depending on the quarks involved in these interactions different elements of the CKM matrix, V_{ij} for particles, and V_{ij}^* for antiparticles, enter into these amplitudes. This means A_f and $\bar{A}_{\bar{f}}$ can have different complex phases, which can result in

$$\left|\frac{A_f}{\bar{A}_{\bar{f}}}\right| \neq 1. \tag{1.49}$$

Thus the decays $\mathrm{P} \to f$ and $\bar{\mathrm{P}} \to \bar{f}$ can have different decay rates - a phenomenon known as 'direct' CP-violation. This can occur for both charged and neutral hadrons, and was

first observed for K^0 decays by the NA48 experiment at CERN and KTeV at Fermilab, at $\mathcal{O}(10^{-6})$, in 1999 [24, 25]. The effect is much larger for B mesons, and was discovered by BaBar and Belle in 2004 using the decay $\mathrm{B}^0_d \to \mathrm{K}^+\pi^-$ [26, 27]. In the absence of any other CP-violating effects one can evaluate the level of direct CP-violation by measuring

$$\mathcal{A}^{\mathrm{CP}} = \frac{N(\mathrm{P} \to f) - N(\bar{\mathrm{P}} \to \bar{f})}{N(\mathrm{P} \to f) + N(\bar{\mathrm{P}} \to \bar{f})}, \tag{1.50}$$

where N simply represents the number of decays of that type observed. As mentioned earlier, LHCb has also recently observed evidence for direct CP-violation in decays of the D^0 by measuring $\Delta\mathcal{A}^{\mathrm{CP}} \equiv \mathcal{A}^{\mathrm{CP}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-) - \mathcal{A}^{\mathrm{CP}}(\mathrm{D}^0 \to \pi^+\pi^-)$ [1].

CP-violation can also occur in a different manner for neutral mesons, but this first requires a description of mixing in neutral mesons.

1.2.3 Mixing in Neutral Mesons and CP-Violation in Mixing

The neutral mesons K^0, D^0, B^0_d and B^0_s consist of a quark and anti-quark that are both down- or up-type but have different flavour. As they are neutral they can transform into their anti-particle via charged weak interactions, as shown for the D^0 in figure 1.2. This causes mixing between the D^0 and $\bar{\mathrm{D}}^0$ in the mass eigenstates, so the mass matrix can be written

$$\begin{pmatrix} M_{\mathrm{D}^0} & \Delta M \\ \Delta M^* & M_{\mathrm{D}^0} \end{pmatrix}, \tag{1.51}$$

for which the eigenvalues are $M_{\mathrm{D}^0} \pm |\Delta M|$. This results in two eigenstates with mass difference $\Delta m \equiv 2\,|\Delta M|$, normally labelled L for 'light' and H for 'heavy'. These can be written in general form as

$$|\mathrm{D}^0_L\rangle = p|\mathrm{D}^0\rangle + q|\bar{\mathrm{D}}^0\rangle \equiv \begin{pmatrix} p \\ q \end{pmatrix}, \tag{1.52}$$

which has mass $M_{\mathrm{D}^0} - \frac{1}{2}\Delta m \equiv m_L$, and

$$|\mathrm{D}^0_H\rangle = p|\mathrm{D}^0\rangle - q|\bar{\mathrm{D}}^0\rangle \equiv \begin{pmatrix} p \\ -q \end{pmatrix}, \tag{1.53}$$

which has mass $M_{\mathrm{D}^0} + \frac{1}{2}\Delta m \equiv m_H$, for some constants p and q such that $|p|^2 + |q|^2 = 1$.

If ΔM is real the solution is $p = q = \frac{1}{\sqrt{2}}$. Using $CP|\mathrm{D}^0\rangle = -|\bar{\mathrm{D}}^0\rangle$ one sees that the mass eigenstates are also CP eigenstates in this case. However, the non-zero complex phases in the CKM matrix allow a non-zero complex phase in ΔM, and so p and q are generally complex. The time evolution of such a state can be described by

$$i\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} p \\ \pm q \end{pmatrix} = \mathcal{H}_{eff}\begin{pmatrix} p \\ \pm q \end{pmatrix}, \tag{1.54}$$

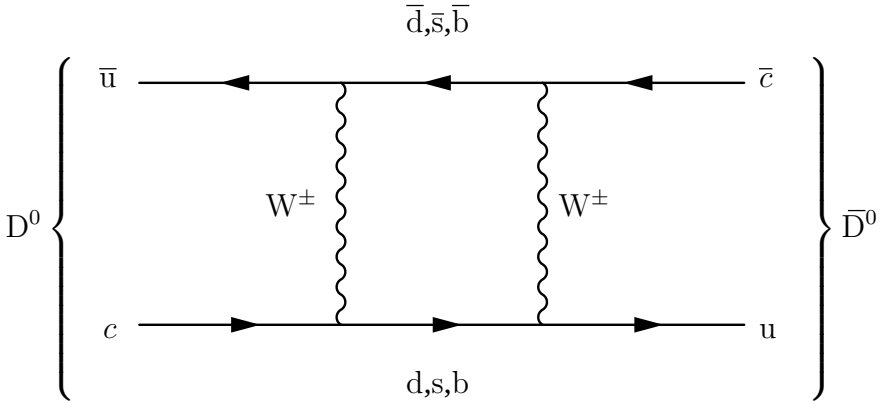


Figure 1.2: One of the dominant diagrams contributing to D^0 mixing.

where the effective Hamiltonian is

$$\begin{aligned}\mathcal{H}_{eff} &= M - \frac{i}{2}\Gamma, \\ &= \begin{pmatrix} \langle \mathrm{D}^0|\mathcal{H}|\mathrm{D}^0\rangle & \langle \overline{\mathrm{D}}^0|\mathcal{H}|\mathrm{D}^0\rangle \\ \langle \overline{\mathrm{D}}^0|\mathcal{H}|\overline{\mathrm{D}}^0\rangle & \langle \mathrm{D}^0|\mathcal{H}|\overline{\mathrm{D}}^0\rangle \end{pmatrix}, \\ &= \begin{pmatrix} M_{\mathrm{D}^0} - \frac{i}{2}\Gamma_{\mathrm{D}^0} & \Delta M - \frac{i}{2}\Gamma_{12} \\ \Delta M^* - \frac{i}{2}\Gamma^*_{12} & M_{\mathrm{D}^0} - \frac{i}{2}\Gamma_{\mathrm{D}^0} \end{pmatrix}. \end{aligned} \tag{1.55}$$

Similarly to the mass matrix, the decay rate matrix Γ has eigenvalues $\Gamma_{\mathrm{D}^0} \pm |\Gamma_{12}|$, thus defining $\Delta\Gamma \equiv 2\,|\Gamma_{12}|$ one has $\Gamma_H = \Gamma_{\mathrm{D}^0} + \frac{1}{2}\Delta\Gamma$ and $\Gamma_L = \Gamma_{\mathrm{D}^0} - \frac{1}{2}\Delta\Gamma$. Consequently, the eigenvalues of $\mathcal{H}_{eff}$ are $\lambda_H = m_H - \frac{i}{2}\Gamma_H$ and $\lambda_L = m_L - \frac{i}{2}\Gamma_L$. The eigenstates are also constrained by

$$\frac{q}{p} = \pm\sqrt{\frac{\Delta M^* - \frac{i}{2}\Gamma^*_{12}}{\Delta M - \frac{i}{2}\Gamma_{12}}} = \pm\frac{2\Delta M^* - i\Gamma^*_{12}}{\Delta m - \frac{i}{2}\Delta\Gamma}, \tag{1.56}$$

where the $\pm$ corresponds to the H and L states respectively.

Solving equation 1.55 gives the time evolution of the mass eigenstates as

$$\begin{aligned}|\mathrm{D}^0_{H,L}(t)\rangle &= e^{-i\lambda_{H,L}t}|\mathrm{D}^0_{H,L}(0)\rangle \\ &= e^{-im_{H,L}t - \frac{1}{2}\Gamma_{H,L}t}\left(p|\mathrm{D}^0(0)\rangle \pm q|\overline{\mathrm{D}}^0(0)\rangle\right). \end{aligned} \tag{1.57}$$

Neutral mesons interact in their flavour eigenstates, and so they are in a pure state of either

$|\mathrm{D}^0\rangle$ or $|\overline{\mathrm{D}}^0\rangle$ at $t = 0$. Using equations 1.52 and 1.53 one has that

$$\begin{aligned}
|\mathrm{D}^0(t)\rangle &= \frac{1}{2p}(|\mathrm{D}^0_H(t)\rangle + |\mathrm{D}^0_L(t)\rangle) \\
&= \frac{1}{2p}\left(e^{-im_Ht-\frac{1}{2}\Gamma_Ht}\left(p|\mathrm{D}^0\rangle - q|\overline{\mathrm{D}}^0\rangle\right) + e^{-im_Lt-\frac{1}{2}\Gamma_Lt}\left(p|\mathrm{D}^0\rangle + q|\overline{\mathrm{D}}^0\rangle\right)\right) \\
&= \frac{1}{2}\left(\left(e^{-im_Ht-\frac{1}{2}\Gamma_Ht} + e^{-im_Lt-\frac{1}{2}\Gamma_Lt}\right)|\mathrm{D}^0\rangle - \frac{q}{p}\left(e^{-im_Ht-\frac{1}{2}\Gamma_Ht} - e^{-im_Lt-\frac{1}{2}\Gamma_Lt}\right)|\overline{\mathrm{D}}^0\rangle\right) \\
&= \frac{1}{2}e^{-iM_{\mathrm{D}^0}t-\frac{1}{2}\Gamma_{\mathrm{D}^0}t}\Bigg(\left(e^{-i\Delta mt-\frac{1}{2}\Delta\Gamma t} + e^{i\Delta mt+\frac{1}{2}\Delta\Gamma t}\right)|\mathrm{D}^0\rangle \\
&\qquad - \frac{q}{p}\left(e^{-i\Delta mt-\frac{1}{2}\Delta\Gamma t} - e^{i\Delta mt+\frac{1}{2}\Delta\Gamma t}\right)|\overline{\mathrm{D}}^0\rangle\Bigg). \qquad (1.58)
\end{aligned}$$

Defining

$$x \equiv \frac{\Delta m}{\Gamma}, \text{ and } y \equiv \frac{\Delta\Gamma}{2\Gamma}, \qquad (1.59)$$

and using $\tau \equiv \Gamma_{\mathrm{D}^0}t$, the probability of the D^0 interacting as a D^0 or $\overline{\mathrm{D}}^0$ at a given time τ after production is

$$P(\mathrm{D}^0 \to \mathrm{D}^0)(\tau) = \frac{1}{2}e^{-\tau}(\cosh(y\tau) + \cos(x\tau)) \qquad (1.60)$$

$$P(\mathrm{D}^0 \to \overline{\mathrm{D}}^0)(\tau) = \frac{1}{2}\left|\frac{q}{p}\right|^2 e^{-\tau}(\cosh(y\tau) - \cos(x\tau)). \qquad (1.61)$$

Performing the same calculation for an initial state of pure $\overline{\mathrm{D}}^0$ one finds that

$$P(\overline{\mathrm{D}}^0 \to \mathrm{D}^0)(\tau) = \frac{1}{2}\left|\frac{p}{q}\right|^2 e^{-\tau}(\cosh(y\tau) - \cos(x\tau)). \qquad (1.62)$$

Thus one sees that if

$$\left|\frac{q}{p}\right| \neq 1, \qquad (1.63)$$

$$P(\mathrm{D}^0 \to \overline{\mathrm{D}}^0)(\tau) \neq P(\overline{\mathrm{D}}^0 \to \mathrm{D}^0)(\tau), \qquad (1.64)$$

which constitutes CP-violation in the mixing rates, known as 'indirect CP-violation'. From equation 1.56 one sees that this is the case if ΔM or Γ_{12} have a complex phase, and a phase difference between them. Although the D^0 system is used here as an example these formulae are generic for any neutral meson system.

This type of CP-violation was the first of any to be discovered, and was observed by Christenson, Cronin, Fitch and Turlay in 1964 in the K^0 system [28]. $\Delta\Gamma$ is very large for the K^0 mass eigenstates, such that the heavy state lives $\sim$100 times longer than the light state. Consequently the states are normally labelled K^0_L, for 'long', and K^0_S, for 'short'. In the absence of CP-violation the K^0_L would have CP eigenvalue -1. The fact that it lives so much longer than the K^0_S allows a pure state of K^0_L to be observed. This was found to decay

to two pions, which is a state with CP eigenvalue +1, showing that $|q/p| \neq 1$ in the K^0 system.

Another type of CP-violation can occur when the final state f is accessible to both the D^0 and $\overline{\mathrm{D}}^0$. Using equation 1.58 and introducing

$$\lambda_f \equiv \frac{q}{p}\frac{\bar{A}_f}{A_f} = -\eta_{CP}\left|\frac{q}{p}\right|\left|\frac{\bar{A}_f}{A_f}\right| e^{i\phi}, \tag{1.65}$$

where $\eta_{CP} = \pm 1$ for CP even or odd final states respectively, and ϕ is the relative phase between q/p and $A_f/\bar{A}_f$, one obtains the instantaneous decay rate of $\overset{(-)}{\mathrm{D}}{}^0(t) \to f$

$$\begin{aligned}\Gamma(\overset{(-)}{\mathrm{D}}{}^0(t) \to f) &= |\langle f|\mathcal{H}|\overset{(-)}{\mathrm{D}}{}^0(t)\rangle|^2 \\ &= \frac{1}{2}e^{-\tau}|\overset{(-)}{A}_f|^2\Big\{\left(1+\left|\lambda_f^{\pm 1}\right|^2\right)\cosh(y\tau) + \left(1-\left|\lambda_f^{\pm 1}\right|^2\right)\cos(x\tau) \\ &\qquad + 2\mathcal{R}(\lambda_f^{\pm 1})\sinh(y\tau) - 2\mathcal{I}(\lambda_f^{\pm 1})\sin(x\tau)\Big\}.\end{aligned} \tag{1.66}$$

Thus one has that $\Gamma(\mathrm{D}^0(t) \to f) \neq \Gamma(\overline{\mathrm{D}}^0(t) \to f)$ if $|\lambda_f| \neq 1$, *i.e.* if either of the criteria for direct or indirect CP-violation are fulfilled, or if $\mathcal{I}(\lambda_f) \neq 0$, *i.e.* $\phi \neq 0$ and $\phi \neq \pi$. The interference between decays of mixed and un-mixed D^0 can cause $\phi \neq 0$ even if CP is conserved in both mixing and decay. Again, this is general for all neutral mesons - the case specific to the D^0 system is discussed in the next section.

1.3 The Charm Sector Parameters y_{CP} and A_Γ

The D^0 system is interesting to analyse as a test of the SM and the CKM mechanism as CP-violation is predicted to be $\mathcal{O}(10^{-3})$ or less. Furthermore, charm physics is the only up-type quark system in which CP-violation can be studied. Significantly greater CP-violation than is predicted by the SM could be achieved through contributions of additional, non-SM particles to the mixing and decay amplitudes. Such particles could include SUSY 'sparticles', or a fourth generation of fermions. Thus high precision CP-violation measurements in the charm sector have great potential to reveal new physics. This makes the aforementioned evidence for direct CP-violation in the D^0 system observed by LHCb particularly exciting. The LHCb detector, described in chapter 2, is well suited for such analyses as it has been specifically designed to perform high precision time-dependent measurements of decays of D and B mesons. Further, the cross section for production of $c\bar{c}$ pairs is very high, $6.1 \pm 0.93\,\mathrm{mb}$ [29], allowing unprecedented numbers of D^0 decays to be recorded.

Mixing in the D^0 system is relatively small and has only been observed in recent years. Both x and y, as defined in equation 1.59, are $\mathcal{O}(10^{-2})$ [3]. This makes it very difficult to

disentangle the heavy and light mass eigenstates. However, one can still obtain sensitivity to the CP violating parameters by measuring the 'effective lifetimes' of the D^0 and $\overline{\mathrm{D}}^0$ [30] - that is, the mean proper decay time of a D that is a D^0 or $\overline{\mathrm{D}}^0$ at production.

Defining A_m and A_d such that

$$\left|\frac{q}{p}\right|^{\pm 2} \equiv 1 \pm A_m, \text{ and } \left|\frac{\bar{A}_f}{A_f}\right|^{\pm 2} \equiv 1 \pm A_d, \tag{1.67}$$

one can use the smallness of x and y to expand equation 1.66 up to second order in τ to give

$$\begin{aligned}\Gamma(\overset{(-)}{\mathrm{D}^0}(t) \to f) \simeq e^{-\tau}|\overset{(-)}{A}_f|^2 \Bigg\{ 1 - \Bigg(& 1 \pm \left(\frac{1}{2}(A_m + A_d) - \frac{1}{4}A_m A_d (A_m + A_d) \right) \\ & - \frac{1}{8}\left((A_m + A_d)^2 + A_m^2 A_d^2 \right) + \frac{1}{2}A_m A_d \Bigg) \\ & \times \eta_{CP}(y\cos\phi \mp x \sin\phi)\tau \\ & + \left[\frac{1}{2}\left(1 \pm \frac{1}{2}(A_m + A_d) + \frac{1}{2}A_m A_d \right) y^2 - \right. \\ & \left. \frac{1}{2}\left(\mp\frac{1}{2}(A_m + A_d) - \frac{1}{2}A_m A_d \right) x^2 \right] \tau^2 \Bigg\}. \end{aligned} \tag{1.68}$$

Naming the coefficient of τ α and that of τ^2 β, the average decay rate is given by

$$\begin{aligned}\hat{\Gamma} &= \Gamma_{\mathrm{D}^0} \frac{1 + \alpha + 2\beta}{1 + 2\alpha + 6\beta} \\ &\simeq \Gamma_{\mathrm{D}^0}(1 - \alpha - 4\beta + 2\alpha^2 + 14\alpha\beta + 24\beta^2). \end{aligned} \tag{1.69}$$

Associating A_d with $\mathcal{O}(10^{-2})$, in accordance with experimental constraints [3], and A_m and $\sin\phi$ with $\mathcal{O}(10^{-1})$, the average decay rates of $\overset{(-)}{\mathrm{D}^0}$ can be written up to $\mathcal{O}(10^{-5})$ as [30]

$$\begin{aligned}\hat{\Gamma}(\overset{(-)}{\mathrm{D}^0} \to f) \simeq \Gamma_{\mathrm{D}^0} \Bigg\{ 1 + & \left[1 \pm \frac{1}{2}(A_m + A_d) - \frac{1}{8}(A_m^2 - 2A_m A_d) \right] \eta_{CP}(y\cos\phi \mp x\sin\phi) \\ & A_m(x^2 + y^2) \pm 2A_m y^2 \cos^2\phi \mp 4xy\cos\phi\sin\phi \Bigg\}. \end{aligned} \tag{1.70}$$

When the D^0 decays to a final state with undefined CP the decay rate is simply the average of the heavy and light decay rates, Γ_{D^0}. Comparing this to the average of the decay rates of D^0 and $\overline{\mathrm{D}}^0$ to a CP-eigenstate final state one obtains the parameter

$$y_{CP} = \frac{\hat{\Gamma}(\mathrm{D}^0 \to f) + \hat{\Gamma}(\overline{\mathrm{D}}^0 \to f)}{2\Gamma_{\mathrm{D}^0}} - 1. \tag{1.71}$$

Using equation 1.70 and expanding up to $\mathcal{O}(10^{-5})$ gives

$$y_{CP} \simeq \eta_{CP} \left\{ \left[1 - \frac{1}{8}(A_m^2 - 2A_m A_d) \right] y\cos\phi - \frac{1}{2}(A_m + A_d)x\sin\phi \right\}, \tag{1.72}$$

and to only $\mathcal{O}(10^{-4})$

$$y_{CP} \simeq \eta_{CP}\left[\left(1+\frac{1}{8}A_m^2\right)y\cos\phi - \frac{1}{2}A_m x\sin\phi\right]. \tag{1.73}$$

Thus, in the absence of CP-violation, $A_m = \phi = 0$, $y_{CP} = \eta_{CP} y$, while any deviation of y_{CP} from y would indicate CP-violation. This makes y_{CP} primarily a measurement of mixing in the D^0 system, requiring an independent measurement of y to test for CP-violation.

Examining the fractional difference of the average decay rates of D^0 and $\overline{\mathrm{D}}^0$ to a CP-eigenstate gives the parameter

$$A_\Gamma = \frac{\hat{\Gamma}(\mathrm{D}^0 \to f) - \hat{\Gamma}(\overline{\mathrm{D}}^0 \to f)}{\hat{\Gamma}(\mathrm{D}^0 \to f) + \hat{\Gamma}(\overline{\mathrm{D}}^0 \to f)}. \tag{1.74}$$

Again, expanding this up to $\mathcal{O}(10^{-5})$ gives

$$\begin{aligned} A_\Gamma \simeq \Bigg[&\frac{1}{2}(A_m + A_d)y\cos\phi - \left(1-\frac{1}{8}A_m^2\right)x\sin\phi - A_m(x^2+y^2) \\ &+ 2A_m y^2\cos^2\phi - 4xy\cos\phi\sin\phi\Bigg]\frac{\eta_{CP}}{1+y_{CP}}, \end{aligned} \tag{1.75}$$

and to $\mathcal{O}(10^{-4})$

$$A_\Gamma \simeq \left[\frac{1}{2}(A_m + A_d)y\cos\phi - x\sin\phi\right]. \tag{1.76}$$

Thus any deviation of A_Γ from zero indicates CP-violation.

If one chooses the CP-even final state $\mathrm{K}^+\mathrm{K}^-$ and the CP undefined state as $\mathrm{K}^-\pi^+$ and its charge conjugate, one can redefine these parameters in terms of the effective lifetimes of the D^0 and $\overline{\mathrm{D}}^0$, using $\tau = 1/\hat{\Gamma}$, as

$$y_{CP} = \frac{\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+) + \tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)}{\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-) + \tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-)} - 1, \tag{1.77}$$

and

$$A_\Gamma = \frac{\tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-) - \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)}{\tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-) + \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)}. \tag{1.78}$$

As y_{CP} only requires measurement of the average of the D^0 and $\overline{\mathrm{D}}^0$ effective lifetimes one need not know the flavour of the D^0 at production. However, for A_Γ, the flavour must be known. For this one can use the decay chain $\mathrm{D}^{*\pm} \to \overset{(-)}{\mathrm{D}}{}^0\pi_s^\pm$, where the charge of the slow pion, $\pi_s^\pm$, determines whether a D^0 or $\overline{\mathrm{D}}^0$ has been produced.

These measurements have already been performed to a high precision at the B factories. BABAR have measured [2]

$$y_{CP} = (11.6 \pm 2.2\,(\text{stat.}) \pm 1.8\,(\text{syst.})) \times 10^{-3}, \tag{1.79}$$

which excludes the no mixing hypothesis at 4.1σ, and BELLE have [4]

$$y_{CP} = (13.1 \pm 3.2\,(\text{stat.}) \pm 2.5\,(\text{syst.})) \times 10^{-3}, \tag{1.80}$$

which is 3.2σ from zero. These are both consistent with the world average of y which, assuming CP-violation is allowed, is [31]

$$y = (8.0 \pm 1.3) \times 10^{-3}. \tag{1.81}$$

Thus, there is still scope to achieve the first single measurement that excludes the zero mixing hypothesis in the D^0 system at a level of 5σ by measuring y_{CP}. Due to the current precision on the measurement of y any CP-violation below $\mathcal{O}(10^{-3})$ will be undetectable at present.

Similarly, BABAR have also measured [32]

$$A_\Gamma = (2.6 \pm 3.6\,(\text{stat.}) \pm 0.8\,(\text{syst.})) \times 10^{-3}, \tag{1.82}$$

and BELLE [4]

$$A_\Gamma = (0.1 \pm 3.0\,(\text{stat.}) \pm 1.5\,(\text{syst.})) \times 10^{-3}, \tag{1.83}$$

both of which are consistent with zero. Direct CP-violation at $\mathcal{O}(10^{-3})$, as observed by LHCb, would contribute to A_Γ at $\mathcal{O}(10^{-4})$ [30], thus A_Γ remains primarily a measurement of CP-violation in mixing. However, the observation of any CP-violation in the charm sector is sufficient motivation to improve the accuracy of the measurement of A_Γ.

Thus, measuring the effective lifetimes of the D^0 and $\overline{\mathrm{D}}^0$ in order to achieve high precision measurements of y_{CP} and A_Γ has high potential for discovery of CP-violation in mixing of the D^0 system, and for confirming mixing at the level of 5σ with a single measurement. This makes them exceptionally interesting parameters to measure at LHCb. The methods used to do so are presented in chapter 4 and the resulting measurements in chapters 5 and 6. These results have been submitted for publication [33].

1.4 Summary

This section provided a review of the mathematical theory behind and the motivation for the measurements presented in chapters 4, 5 and 6. Firstly, in section 1.1 the particle content of the Standard Model was presented, followed by a discussion of the interactions of these particles via the gauge bosons. The Higgs mechanism, and how it bestows masses to the fermions and the $\mathrm{W}^\pm$ and Z^0 bosons was also described. The consequences of there being three generations of fermions, and the mixing between generations that is allowed in their mass eigenstates, was examined. This was elucidated in section 1.2 with a detailed discussion

of the CKM matrix and how the existence of three generations of fermions allows for CP-violation. The phenomenon of mixing in neutral mesons was also introduced, together with that of CP-violation in mixing, and in interference between mixing and decay. Finally, in section 1.3, these phenomena were examined in the specific case of the D^0 system. The parameters y_{CP} and A_Γ, measurements of which are presented in chapters 5 and 6, were introduced. Their implications for discovering CP-violation in mixing in the D^0 system, and confirmation of mixing with a single measurement, were discussed, thus motivating their measurement with increased precision.

Chapter 2

The LHCb Detector

2.1 The LHC Accelerator Complex

The Large Hadron Collider (LHC) [34] is a synchrotron particle accelerator located at the European Organization for Nuclear Research (CERN) [35], on the French-Swiss border, near Geneva. It is housed in a circular tunnel, 27 km in circumference and roughly $100\ \mathrm{m}$ underground, and is part of a chain of accelerators at CERN designed to accelerate and collide bunches of protons. The key figures of merit for such a particle accelerator are: the centre-of-mass energy ($\sqrt{s}$) obtained, as this is the total amount of energy available to create new particles in a single collision; and the instantaneous luminosity ($\mathcal{L}$) delivered, which is the flux (the number crossing a unit of area per unit time) of the circulating particles. The LHC is designed to operate at $\sqrt{s} = 14\ \mathrm{TeV}$, with a bunch crossing rate of 40 MHz, and maximum $\mathcal{L}$ of $10^{34}\ \mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Since starting the commissioning period in September 2009 the LHC has provided collisions at $\sqrt{s} = 7\ \mathrm{TeV}$ throughout 2010 and 2011, operated at a bunch crossing rate of 20 MHz, and obtained a maximum $\mathcal{L}$ of $\sim 3.65 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ [36]. The $\sqrt{s}$ has been increased to 8 TeV for the 2012 run. Following the 2012 run an 18 month shut-down period will commence to allow upgrades to the LHC, so that it can attain its design energy of 14 TeV in the future.

Prior to the injection of the proton bunches into the LHC ring they are passed through a series of older, lower energy accelerators. The full acceleration chain is shown in figure 2.1. The protons start as hydrogen atoms which are stripped of their electrons and passed into the Linear Accelerator (LINAC2), where they reach an energy of 50 MeV; they are then injected into the Proton Synchrotron Booster (PSB) and reach 1.4 GeV before being accelerated to 26 GeV by the Proton Synchrotron (PS). The penultimate acceleration stage is performed by the Super Proton Synchrotron (SPS), where the protons reach 450 GeV, the minimum energy for injection into the LHC. The LHC consists of two beam pipes, each with dipole

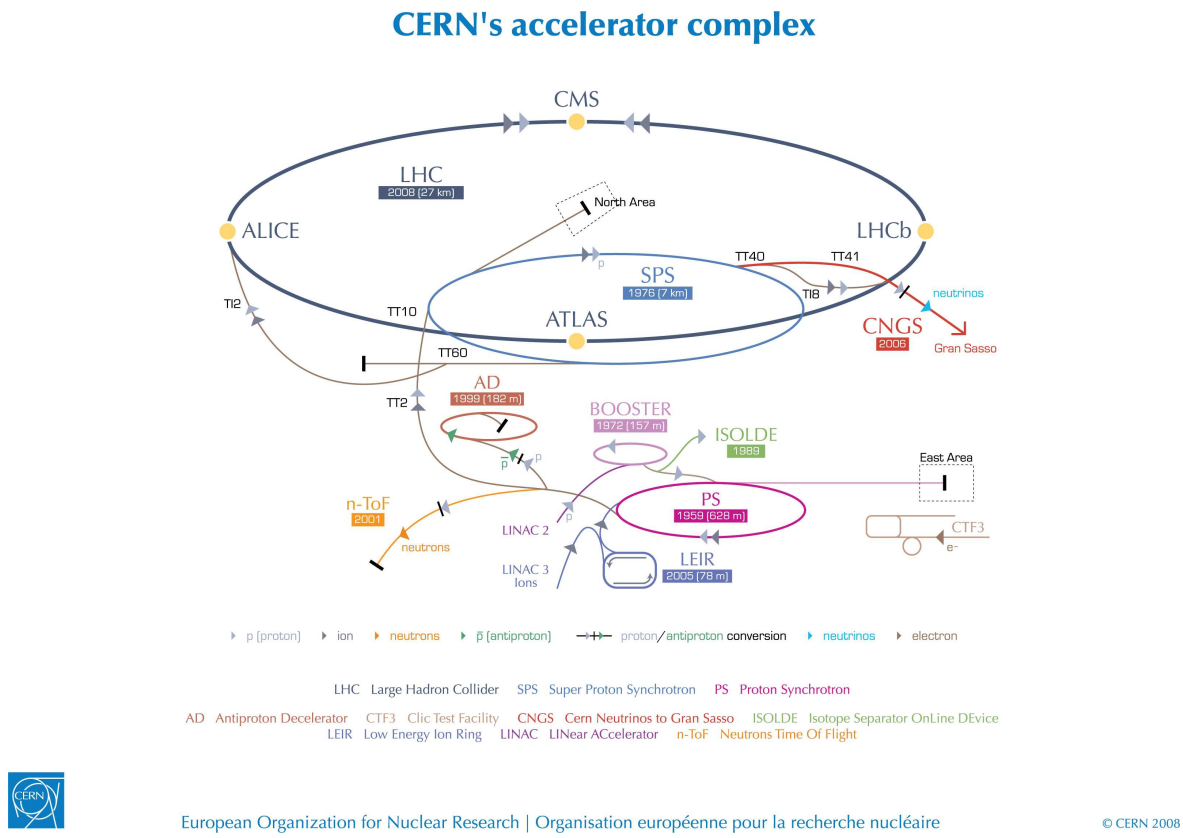


Figure 2.1: The LHC accelerator chain, reproduced from [35].

magnets of opposite polarity, so that two beams of protons can be accelerated in opposite directions. The SPS injects proton bunches into both LHC beam pipes, where each beam is then accelerated to 3.5 TeV (4 TeV in 2012), and collided at $\sqrt{s} = 7$ TeV ($\sqrt{s} = 8$ TeV).

Figure 2.1 also shows the four main LHC experiments, CMS, ATLAS, ALICE and LHCb, situated at the beam crossing points. The Compact Muon Solenoid (CMS) [37] and A Toroidal LHC ApparatuS (ATLAS) [38] are general purpose detectors. Though their designs are somewhat different, they are both barrel shaped detectors that surround the points at which the beams collide, and are intended to find all detectable products of a collision (in the Standard Model (SM) that is all particles except neutrinos). Their main physics programme includes the completion of the SM with the discovery of the Higgs boson (discussed in section 1.1.3), and its extension through the discovery of new, non-SM particles, that might be explained by theories such as Supersymmetry (SUSY). These are generally performed using direct searches, examining invariant mass distributions. They also hope to discover, or place limits on, dark matter candidates and large extra dimensions, by observing unexplained losses of energy in collisions.

The LHC can also accelerate and collide lead nuclei, as was done in November 2010 and 2011 at $\sqrt{s} = 7$ TeV. A Lead Ion Collision Experiment (ALICE) [39] is designed to

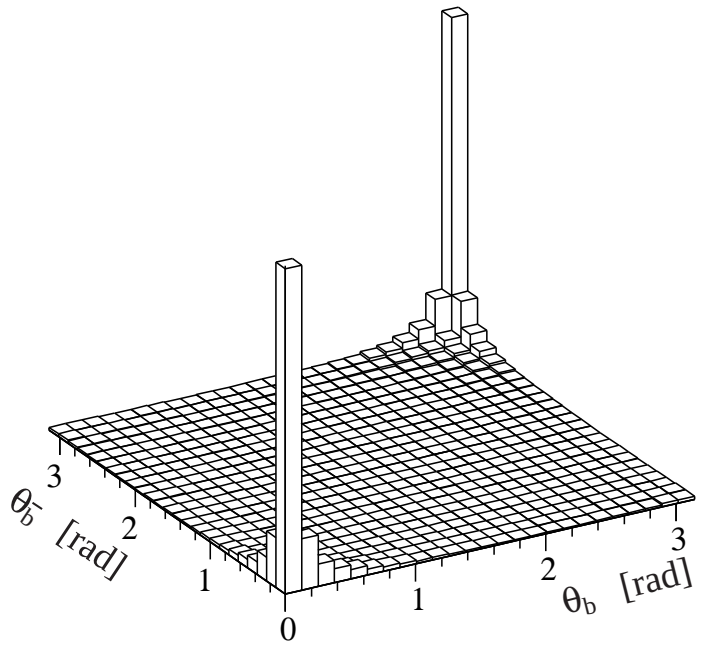


Figure 2.2: The production angles, relative to the beam-line, of $\mathrm{b\bar{b}}$ pairs produced in proton-proton collisions at the LHC, reproduced from [42].

examine such collisions, specifically looking for and examining the nature of the exotic state of matter known as quark-gluon plasma.

The final of the four main LHC experiments is the LHC beauty detector (LHCb) [40]. The data examined in this thesis were taken by LHCb, and so this detector is described in detail in the following section. In addition to these main four there are the TOTEM experiment, which studies p-p interactions, and LHCf, which aims to measure the production cross sections of neutral particles in the forward direction.

2.2 The LHCb Detector

The LHCb detector is designed specifically to observe decays of mesons and baryons containing b and c quarks, and make precision measurements of their properties. b quarks are produced in pairs of b and $\overline{\mathrm{b}}$ in collisions at the LHC. The polar angles, relative to the beam-line, of the b and $\overline{\mathrm{b}}$ are highly correlated, and tend to be very small, as shown in figure 2.2. It is because of this that the LHCb detector has its unique forward-arm design, as shown in figure 2.3. LHCb covers only the region of high pseudo-rapidity, $1.6 < \eta < 4.9$ [41], where $\eta = -\ln(\tan\theta/2)$ and θ is the polar angle from the beam-line. Thus $\sim$30% of b quarks produced in LHC collisions fall within the acceptance of LHCb.

Analyses at LHCb require displaced decay vertices to be reconstructed very cleanly. Consequently, greater signal purity is achieved for events with relatively few reconstructable in-

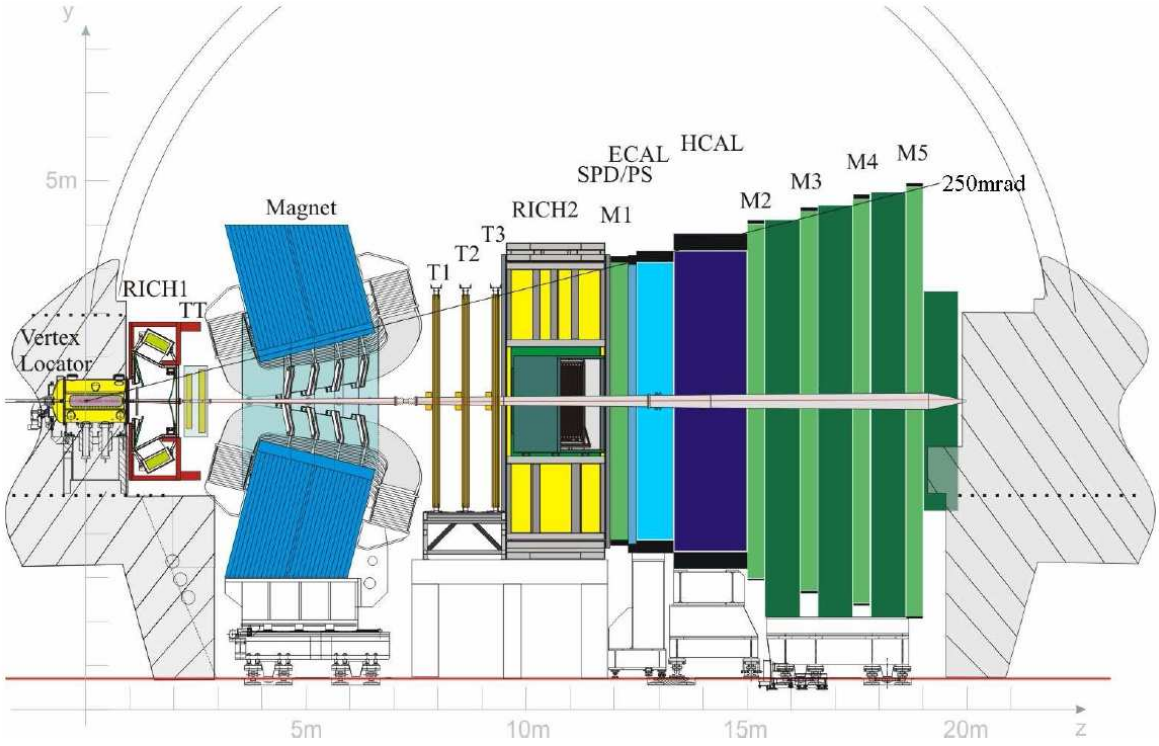


Figure 2.3: The LHCb detector, reproduced from [40].

teractions per bunch crossing. The maximum $\mathcal{L}$ delivered by the LHC to ATLAS and CMS during 2011 resulted in $\mathcal{O}(10)$ interactions per bunch crossing. During this run LHCb operated at $\mathcal{L}$ up to $4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, which provided $\sim$2 reconstructable interactions per bunch crossing. Although this is $\sim$10 times less than the maximum achieved in 2011 it is still considerably larger than LHCb's design luminosity of $2 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. Lower $\mathcal{L}$ is achieved by using less focused beams at LHCb than for ATLAS and CMS. This means the level of focusing can be adjusted as the beams attenuate, so that the $\mathcal{L}$ at LHCb is kept roughly constant. This is shown in figure 2.4, where a comparison to the $\mathcal{L}$ at ATLAS and CMS is made. By the end of the 2011 run the LHC had delivered an integrated luminosity ($\int \mathcal{L}\,\mathrm{d}t$) of 1220 pb^{-1} to LHCb, of which 1107 pb^{-1} was collected by LHCb.

To facilitate a clear frame of reference when discussing the LHCb detector a global coordinate system is defined, and is also shown in figure 2.3. The origin is located at the point at which the two LHC proton beams intersect each other and the protons collide, known as the interaction point. The z-axis is parallel to the line of the proton beams, with positive z pointing into the main LHCb detector, also called the downstream region. The y-axis is in the vertical direction, with positive y pointing upwards, and the x-axis is horizontal, with positive x pointing into the page.

Each sub-detector in LHCb has a specific purpose, intended to allow accurate reconstruction of decays of mesons and baryons containing b and c quarks. In addition, they allow one

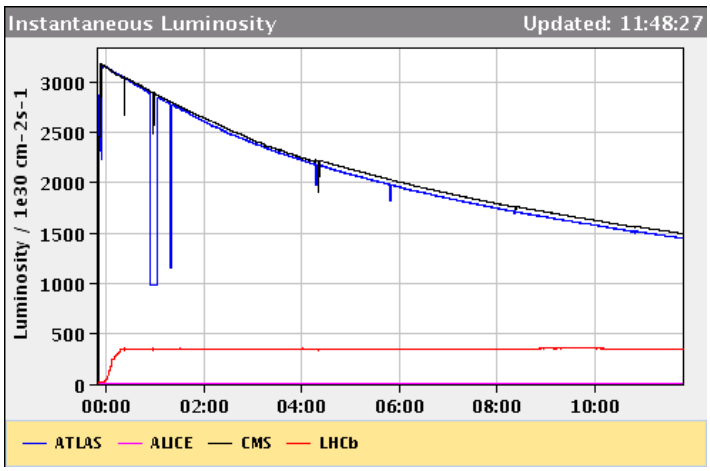


Figure 2.4: Instantaneous luminosity vs time at the LHCb, ATLAS and CMS interaction points, during an LHC fill. Reproduced from [43].

to make as clean a distinction as possible between the decays of interest and the vast number of other particles produced in collisions at LHCb that can fake a signal. Shown in figure 2.3, from left to right, the sub-detectors are: the Vertex Locator (VELO), situated directly around the interaction point, to provide the first tracking points; the first of the Ring Imaging Cherenkov detectors (RICH1), to provide particle identification; the Tracker Turicensis (TT), the first of the large tracking stations; the magnet, to bend the trajectories of charged particles and allow their momenta to be measured; the downstream tracking stations (T1, T2 and T3), to locate charged particles after they have passed the magnet; the second Ring Imaging Cherenkov detector (RICH2); the first of the muon tracking stations (M1), to identify any muons in an event; the electronic and hadronic calorimeters (ECAL and HCAL), to identify electrons, photons, and neutral pions and kaons; and finally four further muon stations (M2 to M5).

The design and performance of each of these sub-detectors will now be discussed in turn, with particular attention paid to the VELO. The performance of the VELO has great influence on the lifetime measurements described in chapters 4, 5 and 6. Its performance in measuring impact parameters is presented in chapter 3.

2.2.1 The Vertex Locator

The Vertex Locator (VELO) [41] is a silicon strip detector that provides precise tracking very close to the interaction point. As its name suggests, it is used to locate the position of any proton-proton collisions within LHCb - known as primary vertices (PVs) - as well as the decay points of any long lived particles produced in the collisions, such as B and D mesons - known as secondary or decay vertices (SVs or DVs). Knowledge of the positions

of the PV and DV, as well as the momentum measurements provided by the downstream tracking stations, allows the calculation of the proper decay time of a long lived particle. This is key to many analyses performed at LHCb, including those presented in this thesis, which study asymmetries between particles and anti-particles as a function of proper decay time. Identifying displaced decay vertices is also of great importance to the LHCb trigger, described in section 2.2.7, which selects which events to keep and which to discard.

2.2.1.1 Design of the VELO

The VELO is required to provide accurate measurements of the trajectories of charged particles, very close to the interaction point. To achieve this it must have very good spatial resolution, and a sufficient number of sensors as to allow the full trajectory to be reconstructed, while keeping the material budget to a minimum. It must also continue to perform well for $\sim$5 years of data taking for LHCb ($\int \mathcal{L} = \sim 8\,\text{fb}^{-1}$), and so must be very radiation hard.

To minimise the extrapolation distance between the first hit of a reconstructed track and the interaction point the active regions of the VELO sensors start at just $8\ \mathrm{mm}$ from the beam-line. This is closer than the minimum safe distance from the beam-line required by the LHC during injection of the beams. For this reason the VELO was constructed in two retractable halves, using 88 roughly semicircular silicon wafer sensors, as is shown in figure 2.5. During injection the VELO is open, with each half retracted from the nominal position by $30\ \mathrm{mm}$, as shown in the bottom right diagram; once stable beams are achieved the VELO is closed, so that the two halves overlap very slightly, as shown in the bottom left diagram. This also requires the VELO to sit inside the beam pipe when fully closed. Consequently, to avoid any degradation of the beam pipe vacuum through out-gassing from the VELO material, the VELO is contained within its own sealed region of vacuum. This is achieved by encasing the two VELO halves in thin aluminium boxes, called the RF-boxes. The inside surface of the RF-boxes is known as the RF-foil. Aluminium was chosen due to its low Z, giving it a relatively low radiation length (X_0). It also serves to insulate the VELO sensors from the electromagnetic field of the beams themselves. The boxes are $300\ \mu\mathrm{m}$ thick at their inner edge, and are corrugated to allow the VELO halves to overlap when fully closed. Figure 2.6 shows one half of the VELO with the RF-box removed.

The sensors are of two types, one measures the radial position (R), and the other measures the azimuthal angle (ϕ) of charged particles that pass through them. The sensors are attached in pairs of R and ϕ types to modules holding the readout electronics. 42 such modules are positioned along the beam-line, 21 on each side; a pair of modules, one on each side of the VELO, is known as a station. The modules in each station are offset from each other by

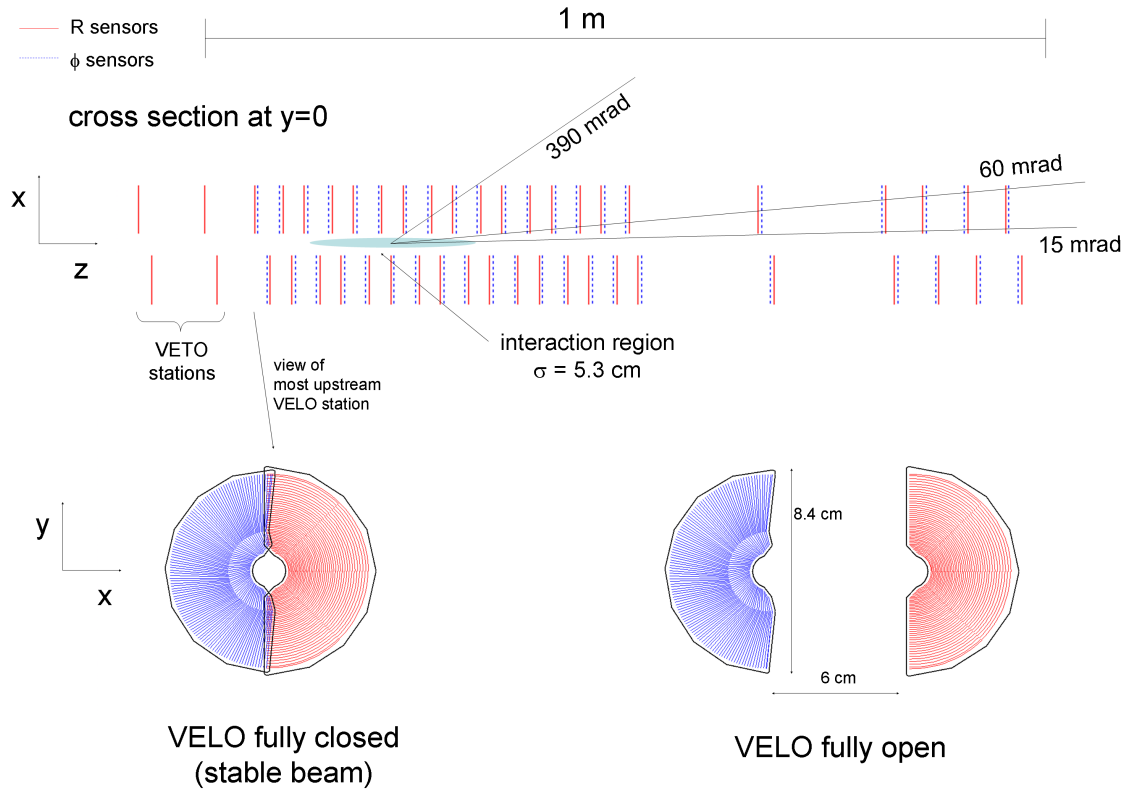


Figure 2.5: The layout of the VELO modules and sensors, reproduced from [41]. The R sensors, which measure the radial position of hits, are shown in red, while ϕ sensors, which measure the azimuthal angle of hits, are shown in blue. R and ϕ sensors are arranged in pairs on modules, which hold the readout electronics.

15 mm, to allow the slight overlap between the two halves of the VELO when it is fully closed. This is to allow full coverage in the R-ϕ plane, and also to aid in the alignment of the modules. The two most upstream stations, labelled 'VETO stations' in figure 2.5, have only R sensors. These were originally intended to be used to make a quick estimate of the number of PVs in an event, so that the first level (L0) hardware trigger could reject events with more than one PV. This idea has since been abandoned, but the VETO stations are still used at L0 to estimate the number of reconstructable tracks in an event, so as to reject overly busy events.

The sensitive region of the sensors themselves consists of a roughly semicircular wafer of silicon 300 $\mu\mathrm{m}$ thick, with an angular coverage of $\sim$182° . The surface of the silicon is covered with aluminium strips, to which a bias voltage is applied. A charged particle passing through the silicon excites electrons which then drift towards the surface of the sensor, and create a measurable signal in the aluminium strips. This is read out to determine the point

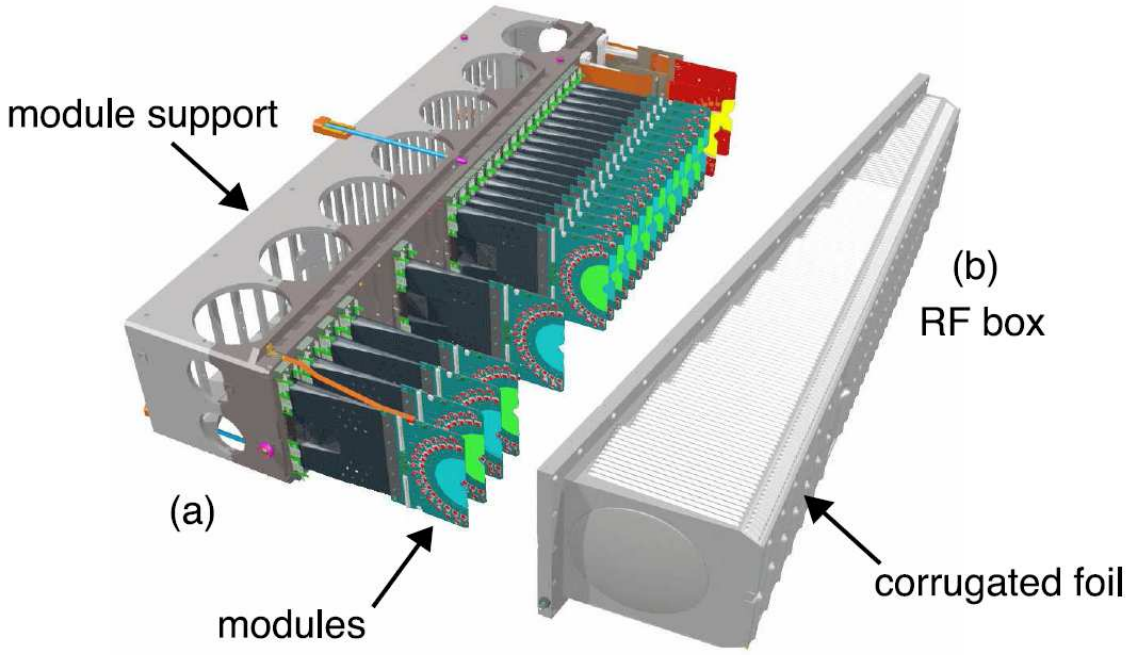


Figure 2.6: One half of the fully assembled VELO with the RF-foil removed, reproduced from [40].

at which the particle intersected the sensor. Each sensor has 2048 strips, making a total of 180,224 readout channels from the VELO. The layout of the strips on each type of sensor is shown in figure 2.7.

The strips on R sensors are circular, so each strip is at a constant radial distance from the centre of the sensor. To minimise the strip occupancy (the fraction of events for which a signal is detected in any single strip), and thus ease track reconstruction, the strips are split along radial lines, spaced 45° apart, into 4 sections of 512 strips each. This shortens the strips, with the shortest closest to the interaction point. This is important as the particle flux varies as $\sim\frac{1}{r^2}$ relative to the beam-line. To further reduce the occupancy, the strips are placed closer together nearer the centre of the sensors than at the outer edges. The pitch (distance between adjacent strips) on R sensors varies from 40 $\mu\mathrm{m}$ at the inside edge to 101.6 $\mu\mathrm{m}$ at the outer edge, while the strip length varies from 3.8 mm to 33.8 mm. This results in an average occupancy of 1.1% for R sensors [40].

The ϕ sensors are designed along similar lines. To reduce the strip length, and thus the occupancy, they are divided into the inner and outer regions, at a radial distance of 17.25 mm. However, the strips do not exactly follow radial lines from the centre of the sensor: in the inner region they make an angle of $\sim$20° to the radial, at the innermost edge; while in the outer region they make an angle of $\sim$-10° to the radial, at 17.25 mm. Furthermore, the ϕ sensors in adjacent modules are flipped, so the inner strips are at $\sim$-20° and the outer strips

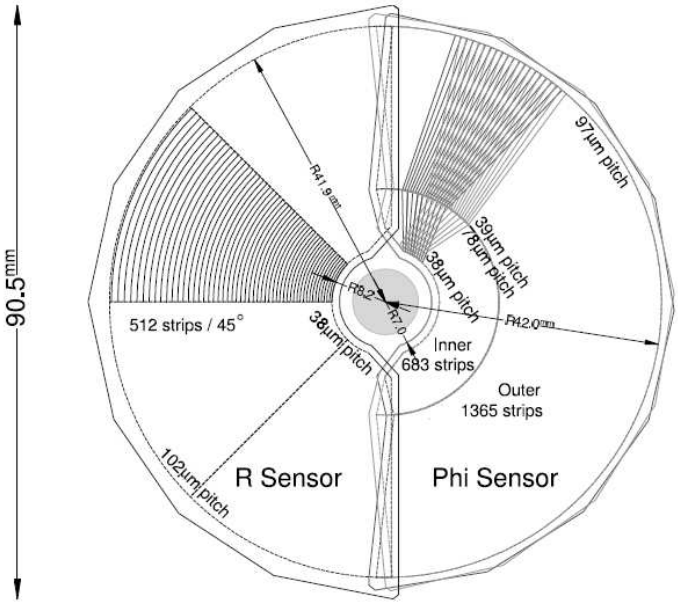


Figure 2.7: The layout of the strips on the R and ϕ sensors of the VELO, reproduced from [40].

at $\sim 10^{\circ}$ to the radial. This is done to further reduce ambiguities in track reconstruction, and hence reduce the frequency with which hits in the VELO are pieced together incorrectly. This layout results in the pitch on a ϕ sensor ranging from 38 $\mu\mathrm{m}$ at the inner edge, to 97 $\mu\mathrm{m}$ at the outer edge, with the strip length ranging from 5.9 mm to 24.9 mm. The average strip occupancy is consequently 1.1% in the inner region, and 0.7% in the outer region [40].

Each VELO sensor uses 16 Beetle readout chips [40] (32 per module), with each chip reading out the charge collected on 128 strips at a rate of 1 MHz. From there the analogue signals are passed to TELL1 readout boards [40], which are based on Field-Programmable Gate Arrays (FPGAs). There are four FPGAs per TELL1 and one TELL1 per sensor. The TELL1 boards perform the basic pre-processing required for the VELO signals. Among these actions is the digitisation of the analogue signals, converting them to integer value analogue-to-digital-converter counts (ADC counts) for each strip. Strips with high ADC counts are then taken as 'cluster' seeds. The strips adjacent to these seeds are added to each cluster, provided their ADC counts are above a certain threshold. The application of a threshold is known as 'zero-suppression' and removes the majority of background noise. The centre-of-charge of each cluster is calculated as a pulse-height weighted position using [44]

$$x_{cluster} = \frac{\Sigma_{strips} ADC_{strip} x_{strip}}{\Sigma_{strips} ADC_{strip}}, \tag{2.1}$$

where x_{strip} is the R or ϕ position of the strip, and ADC_{strip} is the ADC count of the strip. This is taken to be the point at which a charged particle has intersected the sensor. These

data are then sent to data acquisition system (DAQ) for full processing and reconstruction.

The resolution with which the cluster centre is measured depends largely on the spread of the charge deposited in the VELO sensors by a traversing charged particle. If the charge is collected by only one strip one cannot infer anything more about the true position of the hit than the position of that strip - such behaviour is known as 'binary'. If, instead, the charge is shared between two or more adjacent strips, the distribution of the charge across these strips can be used to determine the hit position more precisely [45]. The level of charge sharing between strips is predominantly dependent on the geometry of the strips in the VELO sensors. Finer strip pitch increases the likelihood that charge is shared across several strips. Also, particles intersecting the sensors at larger angles, with respect to the normal to the sensor plane, spread charge more evenly through the silicon. Thus the hit resolution improves with smaller strip pitch, and with larger intersection angle. The critical geometric parameter is the 'projected angle', which is the angle of intersection when projected onto the plane to which the strip is normal.

2.2.1.2 Track and Vertex Reconstruction in the VELO

Clustering is only the first stage of event reconstruction in the VELO. The following stages are performed by the software package MOORE [46], which uses fast reconstruction methods to be used in the trigger, and also by BRUNEL [47], which performs more accurate reconstruction after the data have been taken and stored. R and ϕ clusters on a given module that intersect each other are combined with the known z position of the module to make (R,ϕ,z) space points. A pattern recognition algorithm is then run to select sets of points that may have been created by a single charged particle traversing the detector. These sets of points are then pieced together to make tracks in the VELO. Several different pattern recognition algorithms are run, which can result in the same tracks being found more than once. Consequently, checks are made so that only one instance of each track is kept.

In MOORE, a simple and fast 'line of best fit' calculation is performed to obtain the trajectory for most tracks, while those deemed to be of physics interest are fitted using a fast 'Kalman filter' technique [48, 49]; in BRUNEL a full Kalman filter is performed for all tracks. The Kalman filter finds the best estimate of the track trajectory, taking into account scattering and energy loss of the particle as it passes through the detector material. In all cases, any tracks for which the best fit is not sufficiently good are discarded. This is done to remove so called 'ghost' tracks, that are combinations of random hits and do not correspond to a single charged particle traversing the VELO. Figure 2.8 shows an example of tracks reconstructed by the VELO, from a collision at $\sqrt{s} = 7$ TeV observed during July 2010. Most sensors are hidden so that the clusters and tracks can be seen clearly. Clusters on R sensors are shown in

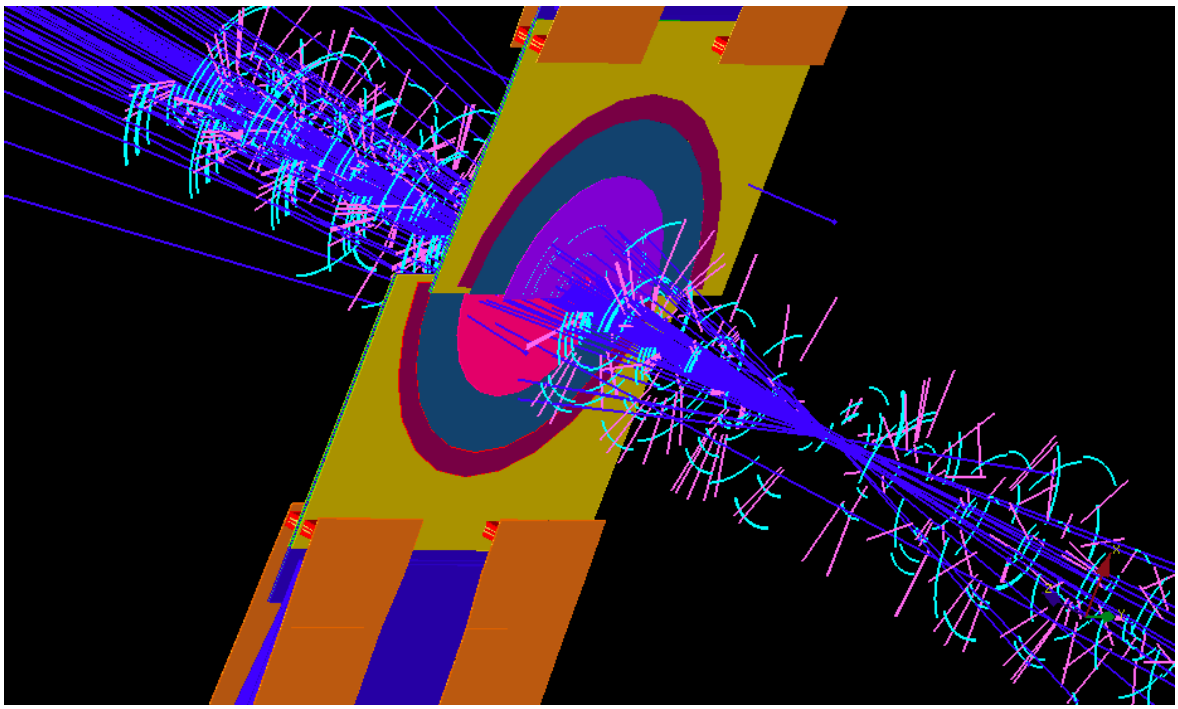


Figure 2.8: A display showing tracks reconstructed in the VELO, during collisions at $\sqrt{s} = 7$ TeV in July 2010, reproduced from [50].

cyan, while clusters on ϕ sensors are in pink; the fitted track trajectories are shown blue.

Further tracking algorithms find track segments in the TT and downstream tracking stations, which are combined with the VELO tracks. Only those tracks which have been reconstructed in both the VELO and TT and/or downstream trackers obtain a momentum measurement, as the particles making them have been displaced by the magnetic field. This will be described in more detail in sections 2.2.3 and 2.2.4.

Once all tracks have been found, the position of any PVs in the event can be determined. MOORE again uses a fast estimation of the PV positions, by simply histogramming the z coordinate of the points of closest approach to the beam-line of all tracks and taking the mean as the z of the PV [48]. BRUNEL uses an adaptive fit method, that iteratively calculates the PV position so as to minimise the χ^2 of the PV, which is calculated as [51]

$$\chi^2_{PV} = \Sigma_{\text{tracks}} \frac{IP^2_{\text{track}}}{\sigma(IP_{\text{track}})^2}, \tag{2.2}$$

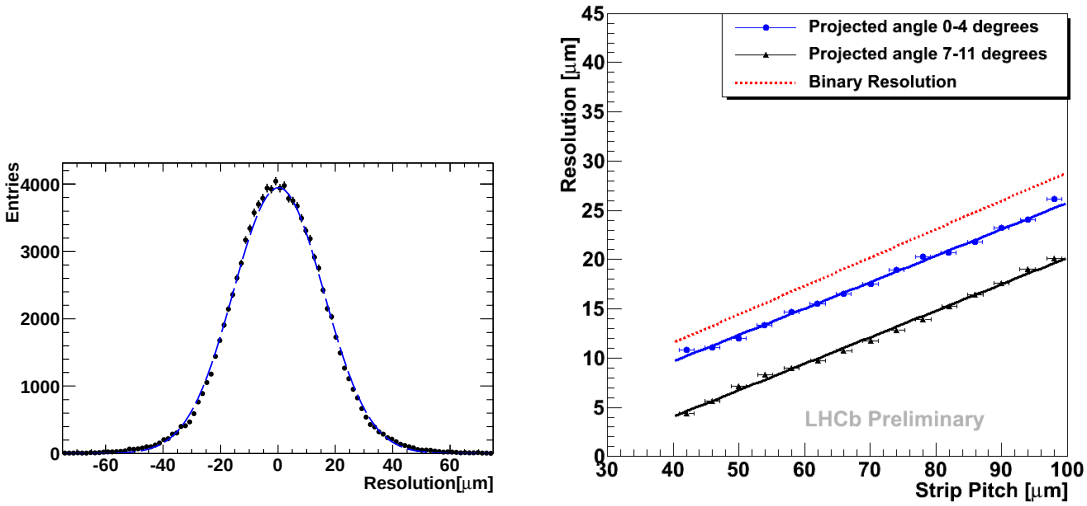
where the impact parameter (IP) of a track is defined as its distance of closest approach to the PV. At each iteration tracks with $IP_{\text{track}}/\sigma(IP_{\text{track}}) > 4$ are excluded from the fit, until convergence is reached. For the best estimate of the error on the IP, $\sigma(IP_{\text{track}})$, the momentum of the track must be known; thus, the PV fit is performed after the VELO tracks have been extended (where possible) into the downstream trackers. The resolution of IP measurements is the topic of chapter 3.

2.2.1.3 Performance of the VELO

The performance of the VELO is of critical importance to the majority of LHCb analyses. The most basic parameter measured by the VELO is the position of single hits, determined by the cluster centres. The worst single hit resolution that can be achieved by a silicon strip detector occurs when electrons excited in the silicon by a traversing particle are only ever collected by a single strip. In this situation, known as 'binary', the single hit resolution is simply given by (strip pitch)/$\sqrt{12}$. The single hit resolution is measured as the σ of a Gaussian fit to the hit residual distributions. Figure 2.9a shows the distribution of residuals in a single bin of projected angle and pitch with a single Gaussian fit, showing that the single Gaussian describes the residuals distribution well. Figure 2.9b shows the resolution of hits on the VELO R sensors, calculated from 2010 data, as a function of strip pitch, in two bins of projected angle, with a comparison to the binary situation. The best resolution is 4 $\mu\mathrm{m}$ for 40 $\mu\mathrm{m}$ pitch and large projected angles. This is the best single hit resolution achieved by any LHC detector. As expected, the resolution increases with larger pitch and lower projected angles. The resolution could potentially be further improved, particularly for tracks at low projected angles, by studying the distribution of charge within clusters. By doing so one could modify equation 2.1 to account for any non-linear dependence of the charge sharing on the true point of intersection, as well as the variation in the level of charge sharing as a function of strip pitch and projected angle [45].

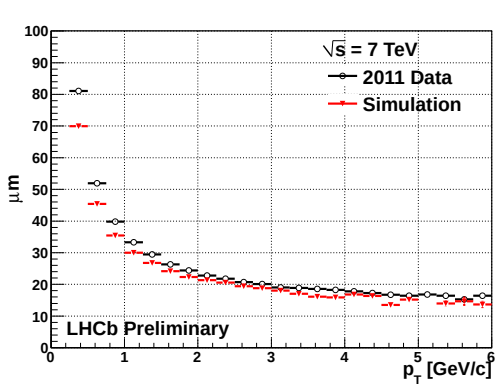
The resolution on the positions of PVs found in the VELO is also of great importance. Equation 2.2 shows that the driving factor in this is the resolution on IP measurements. As discussed in chapter 3 the IP resolution improves with increasing particle momentum. Figure 2.10a shows the resolution on the x component of IP (IP$_x$) measurements as a function of p_T. Again, the performance of the VELO is very good, achieving a resolution on IP$_x$ of < 36 $\mu\mathrm{m}$ for particles with $p_T > 1\,\mathrm{GeV}$. This is in reasonable agreement with expectations from simulation. The excellent IP resolution is reflected in the PV resolution. The resolutions on the x and y PV co-ordinates are shown, in figure 2.10b, as a function of the number of tracks included in the PV fit. For a PV using 25 tracks in its fit the resolution on the x coordinate of its position is just 13.1 $\mu\mathrm{m}$, while the resolution on the y coordinate is just 12.5 $\mu\mathrm{m}$ [50]. The average PV reconstructed in LHCb in fact uses > 50 tracks, but the method of measuring their resolution requires dividing the constituent tracks of a PV into two sets and fitting a PV with each set.

Thus the VELO has been found to perform exceptionally well in all key areas for physics measurements. Its performance with respect to IP measurements is discussed in detail in chapter 3.

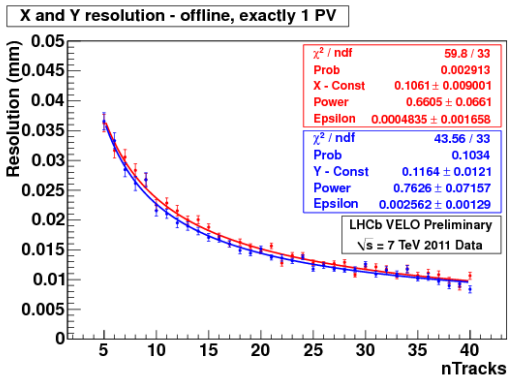


(a) **(b)**

Figure 2.9: (a) The single hit residuals for hits with projected angle $< 4\ ^\circ$ and pitch between 60 μm and 64 μm, fitted with a single Gaussian. (b) The resolution of R hits in the VELO, as a function of strip pitch, in bins of projected angle, with comparison to the resolution of a binary detector. Reproduced from [50].



(a) Resolution of the x component of IP measurements (IP_x) as a function of p_T, comparing data taken during 2011 to simulated data.



(b) Resolution of PV x and y coordinates, reproduced from [50].

Figure 2.10

2.2.2 The Ring Imaging Cherenkov Detectors

The two Ring Imaging Cherenkov (RICH) detectors at LHCb provide particle identification for the experiment - that is, determination of the species of particles observed in the detector (*i.e.* electrons (e), muons (μ), pions (π), kaons (K), or protons (p)). To do this they exploit the phenomenon of Cherenkov radiation, whereby a charged particle traversing a di-electric material (radiator) at a velocity greater than the local speed of light in that material emits photons. These photons are produced at an angle to the particle's trajectory that is dependent on its velocity, v:

$$\cos\theta_C = \frac{c}{nv}, \tag{2.3}$$

where c is the speed of light in vacuum, and n is the refractive index of the material [52]. They are thus emitted in a cone around the particle as it traverses the material, and can be observed as a ring when they intersect a photo-sensitive surface. For a given p each different species of charged particle will produce a ring with a different radius. Thus knowing the p of a given track, one can compare the expected rings with the photons observed and so infer the species of the particle that made the track.

2.2.2.1 Design of the RICH Detectors

There are two RICH detectors in LHCb as shown in figure 2.3: RICH1 [53] is positioned before the magnet and is designed to perform particle identification (PID) for low momentum particles; RICH2 [54] is situated downstream of the magnet, and is designed to perform PID for high momentum particles. The momentum range covered depends on the radiator material used: RICH1 uses aerogel, with $n \simeq 1.03$, and $\mathrm{C_4F_{10}}$ gas, with $n \simeq 1.0014$; while RICH2 uses $\mathrm{CF_4}$ gas, with $n \simeq 1.0005$. Figure 2.11 shows the dependence of θ_C on particle momentum for the different radiators and species of particle. At high momentum all particle species become indistinguishable by their θ_C - this is known as the 'saturation' point. Thus RICH1 covers particles with momenta in the range $\sim$1 GeV to $\sim$70 GeV , while RICH2 covers the range from $\sim$15 GeV to $>$ 100 GeV. RICH1 also has a much larger angular coverage, so as to provide PID for all particles within the acceptance of the downstream tracking stations. Its acceptance starts at 25 mrad and extends to 300 mrad in the horizontal axis and 250 mrad in the vertical axis, while RICH2 has a smaller coverage, from 15 mrad to 120 mrad in the horizontal and 100 mrad in the vertical. Figure 2.12 shows schematics of RICH1 and RICH2.

To observe the rings the Cherenkov photons must be collected; this is done using arrays of Hybrid Photon Detectors (HPDs). A schematic of an HPD is shown in figure 2.13. An HPD is a vacuum tube with a quartz window, coated with a photo-cathode material, at the

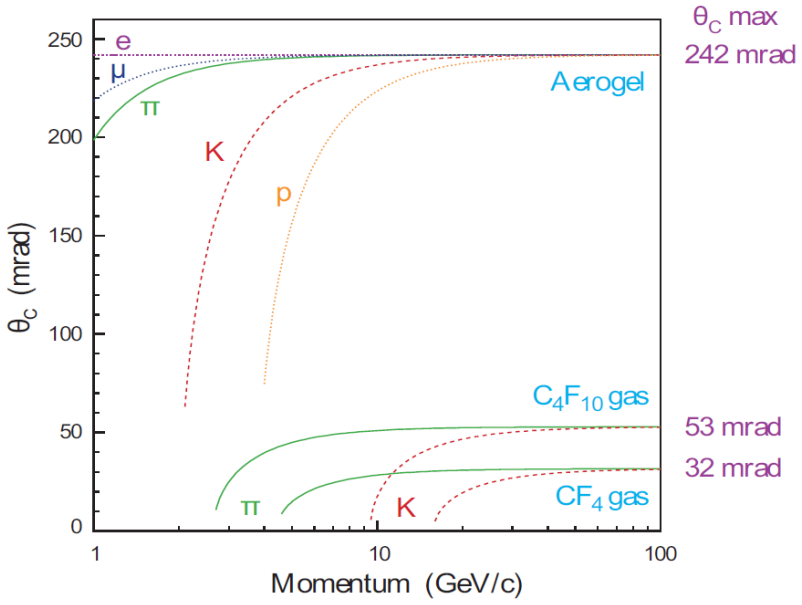
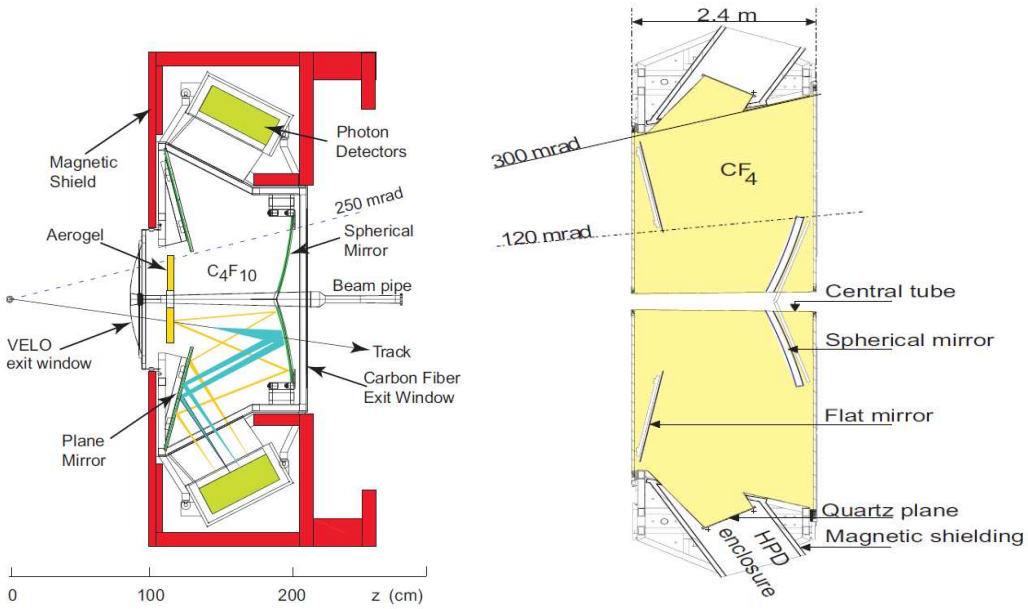


Figure 2.11: The Cherenkov angle, θ_C, of photons produced by different particles in different radiators, as a function of the particle momentum. Reproduced from [40].



(a) The RICH1 detector, in the y-z plane.

(b) The RICH2 detector, in the x-z plane (top view).

Figure 2.12: Schematics of the RICH detectors, reproduced from [40].

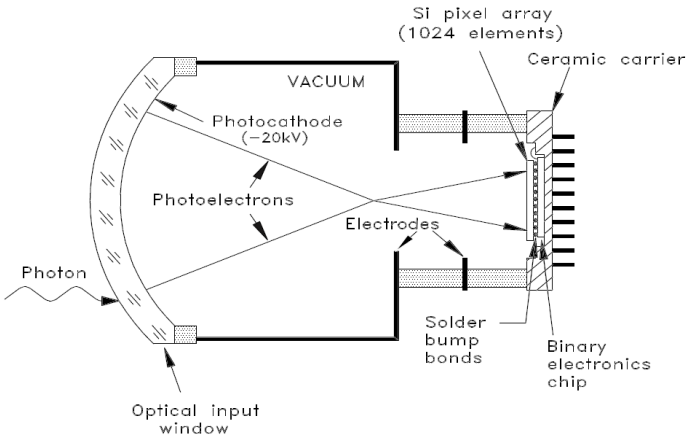


Figure 2.13: A schematic of a Hybrid Photon Detector (HPD) as used in the RICH detectors. Reproduced from [40].

detection end. Photons incident on this window produce photo-electrons which are then accelerated through a potential of $\sim$20 kV, to be collected by an array of silicon pixels at the readout end of the tube. This gives the HPDs the ability to detect single photons with a good efficiency.

The HPDs are carefully shielded from the magnetic field of the LHCb dipole, but a small field of a few mT still penetrates to the HPDs. This modifies the trajectories of the photo-electrons within them from a simple straight line. Consequently, inferring the point at which the Cherenkov photon was incident on the quartz window is non-trivial, and care must be taken to account for the magnetic field. Calibration is achieved by shining known patterns of photons onto the HPDs and monitoring how the hit pattern on the silicon pixels is modified by a non-zero magnetic field.

As the HPDs are necessarily very sensitive they must be mounted outside the path of the majority of particles produced in collisions, to minimise background noise and avoid damage to their silicon pixels. Arrays of highly reflective mirrors are thus used to direct the Cherenkov photons into the HPD arrays - their positions in RICH1 and RICH2 are shown in figure 2.12. The first set of mirrors in both RICH detectors are spherical, and serve to focus the rings of photons and direct them out of the LHCb acceptance; the second set are flat and simply direct the photons into the HPD arrays. The mirrors must be very precisely aligned and monitored for movement - this is done via the Laser Alignment Monitoring System (LAMS) [55].

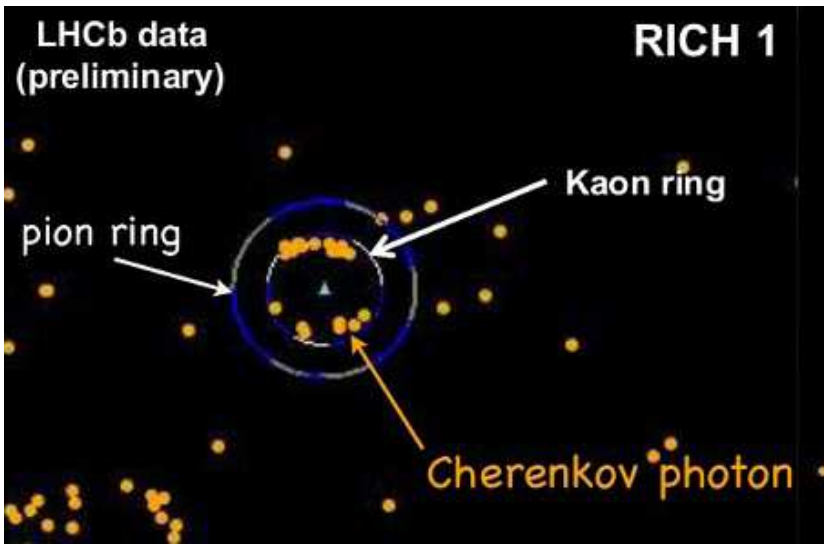


Figure 2.14: An example of the distribution of photons observed in the HPDs of the RICH1 detector, with a comparison to the Cherenkov rings expected for a given track under the π and K hypotheses. Reproduced from [57].

2.2.2.2 Particle Identification in the RICH Detectors and its Performance.

Each track reconstructed in LHCb is extrapolated to the HPD arrays as if it had been reflected by the RICH mirrors; its point of intersection with the HPDs then lies at the centre of any rings of Cherenkov photons it may have produced. The PID algorithm tries five mass hypotheses for the track (e, μ, π, K, and p). For each mass hypothesis it uses the momentum measurements provided by the tracking systems to calculate the expected θ_C for that track, and thus the expected radius of any Cherenkov rings produced, should it be of that species. By comparing the hypothesised ring radius with the distribution of the photons observed it constructs a likelihood for each mass hypothesis [56]. Figure 2.14 shows an example of the distribution of photons observed in the HPDs of RICH1 compared with the expected distribution for a given track under the π and K hypotheses. There is a clear ring observed that matches the K hypothesis, while only one photon hit lies on the ring expected from the π hypothesis. Thus this track is assigned a high likelihood of being a K.

Such a process has a certain rate of mis-identification, whereby it identifies a track as being of a certain species other than its true identity. The efficiency with which the PID algorithm performs can be checked, without the use of simulation, by using decay channels for which the kinematics of the decay products are sufficient to identify them without using the RICH detectors. To check this for π-K separation the decay $\mathrm{D}^{*+} \rightarrow \mathrm{D}^0(\mathrm{K}\pi)\pi^+$ is used, as applying a tight constraint on $m(\mathrm{D}^{*+}) - m(\mathrm{D}^0)$ is sufficient to select a very clean signal sample. Figure 2.15 shows the efficiency, as a function of momentum, of correctly identifying a K as a K, and wrongly identifying a π as a K. As expected, correct PID

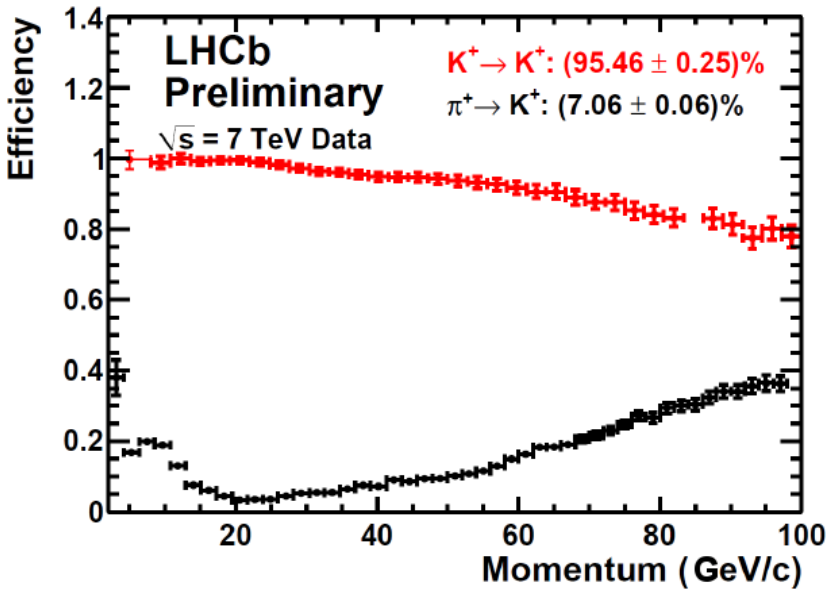


Figure 2.15: The efficiency, as a function of particle momentum, with which the RICH detectors correctly identify a K as a K, and wrongly identify a π as a K. Measured on 2010 data. Reproduced from [57].

becomes much more difficult at high momentum. The mis-ID rate at low momentum is also higher than was expected. This is because the aerogel, which covers the low momentum region, absorbed the $\mathrm{C_4F_{10}}$, somewhat reducing its efficacy as a radiator. However, the performance of the other radiators has been as expected from simulation. Thus, the RICH detectors perform their task very well, achieving on average a correct PID rate of > 90 % and a mis-ID rate of < 10 % for π, K and p. They are thus indispensable for most physics analyses at LHCb.

2.2.3 The Dipole Magnet

The dipole magnet at LHCb [58] provides an integrated magnetic field of $\sim$4 Tm in order to displace the trajectories of charged particles and allow their momenta to be measured. A diagram of the magnet is shown in 2.16. It is a warm (not super-conducting) magnet consisting of two identical, saddle shaped aluminium conducting coils positioned symmetrically above and below the beam-line. It operates at a nominal current of 5.85 kA. Its polarity can readily be reversed, so as to cancel any asymmetries in the detection efficiency that might fake CP-violation. Throughout data-taking in 2010 and 2011 this has been done regularly, and an approximately equal quantity of data has been taken with each polarity.

In order to make accurate momentum measurements the magnetic field strength throughout the detector must be known very precisely. To this end a survey of the magnetic field

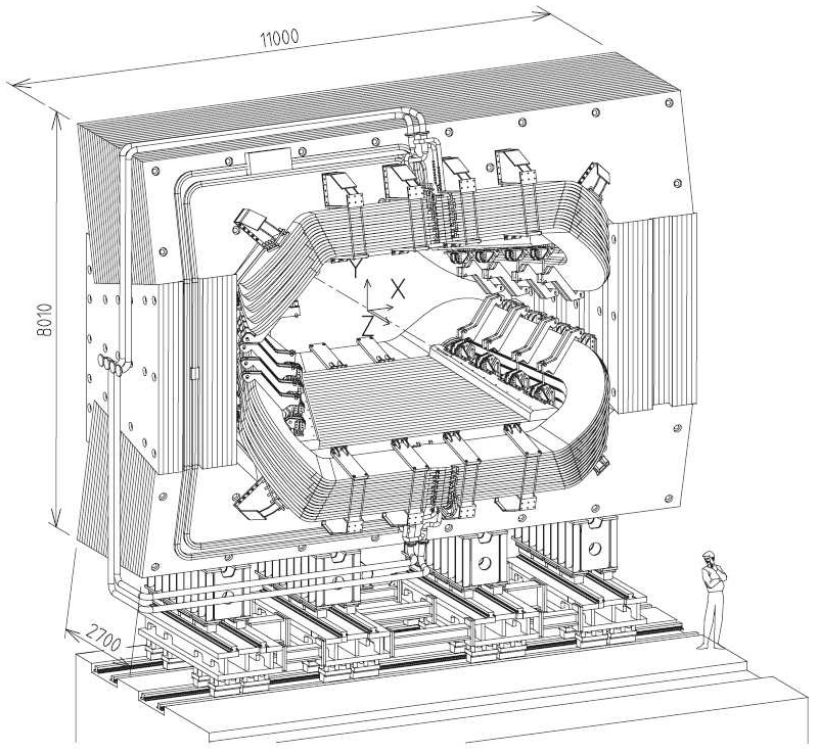
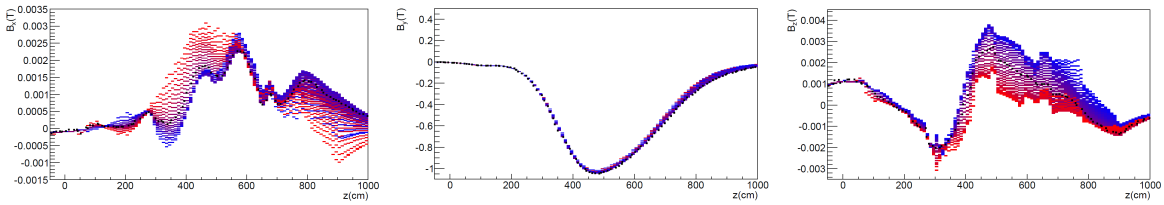


Figure 2.16: A diagram of the LHCb magnet, reproduced from [40].

strength was made in 2005 and again in 2011 using Hall probes. This achieved a spatial accuracy of 0.2 mm and a magnetic field strength accuracy of $\sim$0.2 mT . The results of the 2011 survey are shown in figure 2.17. Over a distance of $\sim$5 m it provides an average magnetic field strength of $\sim$0.8 T in the y direction, thus achieving an integrated field of $\sim$4 Tm . The resolution of momentum measurements achieved is discussed in section 2.2.4.

2.2.4 The Tracking System

The tracking stations downstream of the VELO serve to provide measurements of the trajectories of charged particles before and after the magnet, to allow measurement of their momenta. There are four stations: the TT, positioned before the magnet, and T1, T2 and T3 downstream of the magnet. The TT and the inner regions of T1-T3 are subject to very high particle flux, thus they must be very radiation hard, and have sufficiently high granularity as to keep occupancies low enough for reliable pattern recognition. For these reasons they consist of silicon strip sensors. They are collectively referred to as the Silicon Tracker (ST), with the inner regions of T1-T3 alone known as the Inner Tracker (IT). The outer regions of T1-T3 - known as the Outer Tracker (OT) - suffer significantly less irradiation, and so cheaper ‘straw tube’ drift-time sensors are used.



(a) **(b)** **(c)**

Figure 2.17: The results of the 2011 magnetic field strength survey. (a) shows B_x, (b) B_y, and (c) B_z as a function of z coordinate. The measurements are taken at $y = 0$. The data-points in blue are at positive x, those in red at negative x, and those in black at $x = 0$. Reproduced from [59].

2.2.4.1 The Silicon Tracker

The Tracker Turicensis (TT) [40, 60] is a planar tracking station, 150 cm wide and 130 cm high, covering the full angular acceptance of LHCb. It is situated after RICH1 and before the magnet, as shown in figure 2.3. It consists of four planes of silicon strip sensors with an (x-u-v-x) layout, shown in figure 2.18; the x layers have their detection strips aligned vertically, while the u and v layers have theirs rotated through -5° and $+5^\circ$ to the vertical respectively. This is done for similar reasons as the displacement of the strips on the VELO ϕ sensors from the radial line, as described in section 2.2.1.1 - that is, to remove ambiguities in hit association and ease pattern recognition. The TT silicon sensors are 500 $\mu\mathrm{m}$ thick, 9.6 cm long and 9.4 cm wide; they have a strip pitch of 183 $\mu\mathrm{m}$ and 512 readout channels each, giving it a total active area of 8.4 m^2 and 143,360 readout channels.

The Inner Tacker (IT) [40, 62] makes up the inner region of the three downstream tracking stations T1-T3 . The sensors are arranged in a cross shape, 120 cm wide and 40 cm high, about the beam-pipe, as shown in figure 2.19. Each station has four layers with the same (x-u-v-x) layout as the TT. The IT sensors are either 320 $\mu\mathrm{m}$ or 410 $\mu\mathrm{m}$ thick and are 7.6 cm wide and 11 cm long; they have a strip pitch of 198 $\mu\mathrm{m}$ and 384 readout channels each. This gives the IT an active area of 4.0 m^2 and a total of 129,024 readout channels.

2.2.4.2 The Outer Tracker

The Outer Tracker (OT) [40, 63] makes up the portions of T1-T3 farthest from the beam-line. Due to the lower particle flux in this region cheaper ‘straw-tube’ drift-time sensors are used. Each ‘straw-tube’ consists of an outer tube with a 5 mm diameter, made from 25 $\mu\mathrm{m}$ thick polyimide, to make them gas tight, and 12.5 $\mu\mathrm{m}$ aluminium, to transmit the signal and provide electrical shielding. At the centre of each tube is a cathode of 40 $\mu\mathrm{m}$ thick carbon doped polyimide. The tubes contain a mixture of 70 % Argon and 30 % CO_2 which

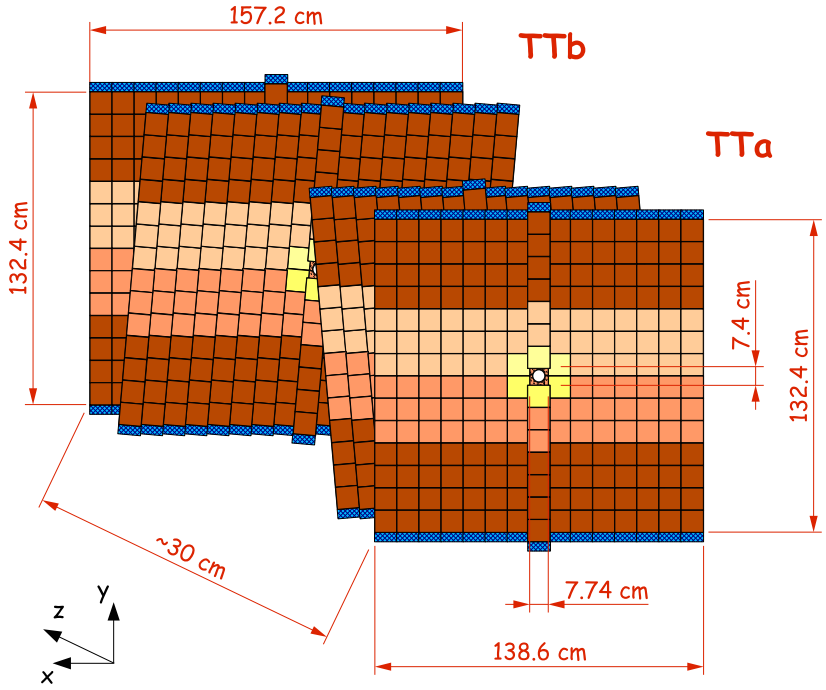


Figure 2.18: The layout of the four detection layers of the TT. Reproduced from [61].

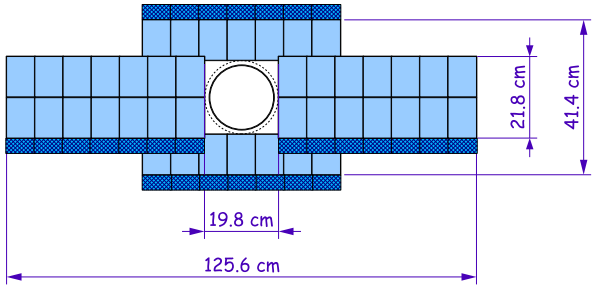


Figure 2.19: The layout of the IT sensors. Reproduced from [61].

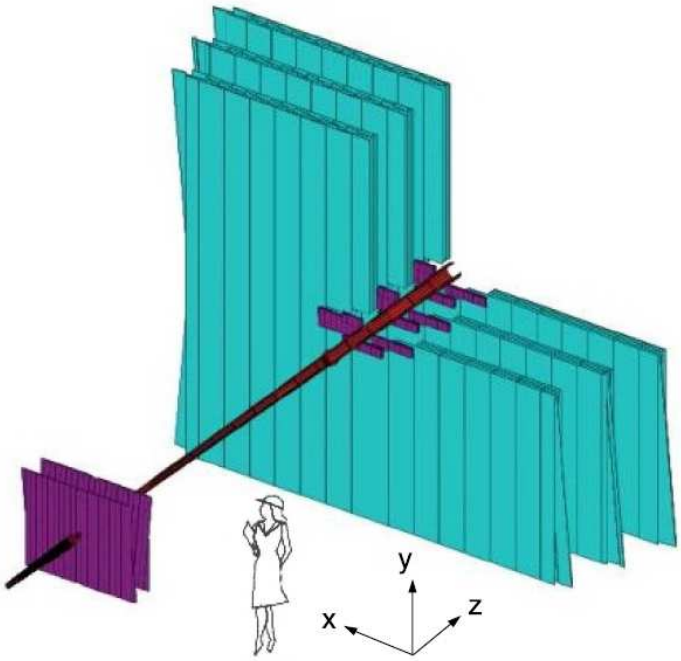


Figure 2.20: The layout of the TT, IT, and OT together. The TT and IT are shown in purple, and the OT in blue, with the beam-pipe shown in red. Reproduced from [40].

is ionised by traversing charged particles; the electric field of the inner cathode attracts the ionisation electrons which create a measurable signal. The chosen gas mixture ensures that the electron drift time is < 50 ns for the majority of hits.

Each module comprises two staggered layers of 64 tubes, making 128 readout channels; the diameter of the tubes give the sensors a pitch of $\sim$5 mm. The three OT stations consist of four layers of sensor modules, arranged in the same (x-u-v-x) layout as used in the ST. The stations have acceptance up to 300 mrad in the horizontal axis, and 250 mrad in the vertical; each station has an active area of $\sim$29 m^2 , giving a total of $\sim$55,000 readout channels. The layout of the TT, IT, and OT together is shown in figure 2.20.

2.2.4.3 Performance of the Tracking System

Reconstruction in the tracking stations follows much the same procedure as was described for the VELO in section 2.2.1.2: pattern recognition, followed by duplicate removal and track fitting. The track segments found in the VELO are then combined with those in the TT and T1-T3 . The knowledge of the magnetic field, described in section 2.2.3, is used to estimate the curvature of the particle trajectories as they traverse the detector and thus provide a momentum measurement.

The ST has achieved a hit resolution of $\sim$58 μm , while the OT has a hit resolution of $\sim$230 μm [64]. Further, the full tracking system has achieved its target momentum

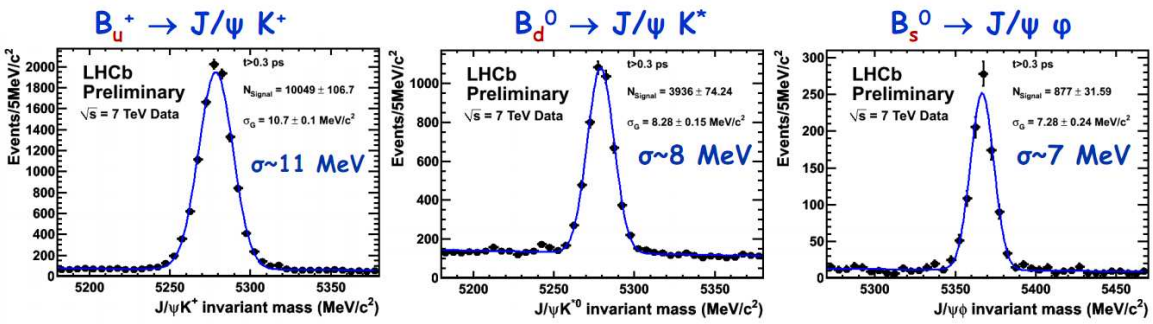


Figure 2.21: An example of the invariant mass resolutions achieved by the LHCb tracking system using 2010 data. Reproduced from [64].

resolution of $\sigma(p)/p \simeq 0.4\%$, and consequently achieved some exceptional mass resolutions, as shown in figure 2.21.

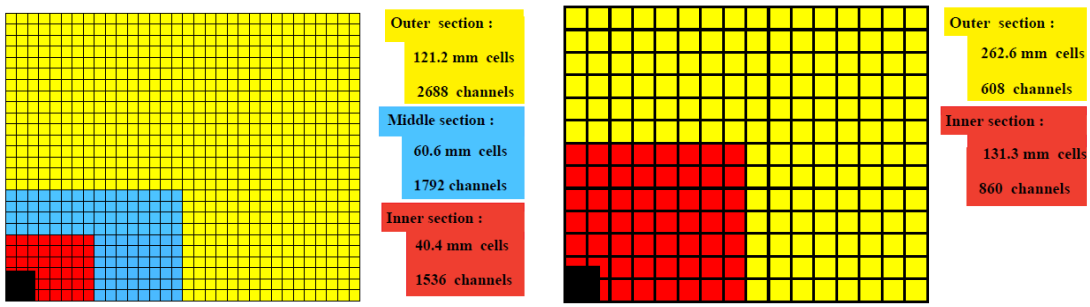
2.2.5 The Calorimeters

LHCb has two calorimeters [40, 65], situated between the first and second muon tracking stations: the Electromagnetic Calorimeter (ECAL) is used to provide position and energy measurements for photons and electrons; the Hadronic Calorimeter (HCAL) does similarly for hadrons - neutral hadrons in particular, as they cannot be detected by the LHCb tracking system. They are particularly useful in the first level (L0) trigger, as is described in section 2.2.7, as well as for PID. In order to distinguish electrons from hadrons the ECAL is preceded by a Scintillator Pad Detector (SPD), that identifies charged particles and allows rejection of neutral hadrons, and a Pre-shower detector (PS), which detects electromagnetic showers, and thus identifies only electrons.

2.2.5.1 Design of the Scintillator Pad and Pre-shower Detectors

The SPD and PS detector modules use pads that scintillate when traversed by charged particles. The light thus produced is read out by wavelength-shifting (WLS) fibres and then reflected along clear plastic fibres to photo-multiplier tubes (PMTs). In the case of the SPD and PS multi-anode PMTs (MaPMTs) are used - they receive 64 fibres each, and allow the fibres to be read out individually, giving a fine granularity to the sensors.

The SPD and PS are in fact almost identical in their design, except for the fact that the components of the SPD are all ∼0.45 % smaller than those of the PS. They consist of rectangular scintillator pads of varying size, depending on their distance from the beam-line, as shown in figure 2.22a. This is to account for the vast difference in occupancy between the inner and outer edges of the detector, as has been discussed for the VELO and tracking



(a) The segmentation of the SPD, PS, and ECAL sensors.

(b) The segmentation of the HCAL sensors.

Figure 2.22: Diagrams showing the segmentation of the calorimeter components. Reproduced from [40].

stations. Their active area is 7.6 m wide and 6.2 m high, and they have a total of 12,032 readout channels.

They are placed on either side of a 15 mm lead plate, which corresponds to 2.5 X_0, and causes electrons to produce electromagnetic showers. Thus the SPD determines whether a traversing particle is charged, then the PS determines whether it has created an electromagnetic shower in the lead plate, allowing the identification of electrons.

2.2.5.2 The Design of the Electromagnetic Calorimeter

The ECAL is designed to provide position and energy measurements for electrons and photons. Each ECAL module consists of 2 mm of lead, to induce electromagnetic showers, followed by 4 mm of scintillator material. The scintillator pads have a similar design to the SPD and PS sensors, using WLS fibres to read out the light produced, except that the fibres are grouped in bundles, and each bundle is passed to a single PMT, giving a coarser granularity. The ECAL consists of 66 layers of such modules, with each layer arranged as shown in figure 2.22a. This results in a total depth of 42 cm, which corresponds to 25 X_0. The energy resolution thus achieved is

$$\frac{\sigma_E}{E} = \frac{\sim 9\%}{\sqrt{E}} \oplus \sim 0.8\%, \tag{2.4}$$

where E is in GeV.

2.2.5.3 The Design of the Hadronic Calorimeter

The HCAL is designed to provide position and energy measurements for hadrons, particularly neutral hadrons that cannot be detected by the tracking stations. Similarly to the ECAL

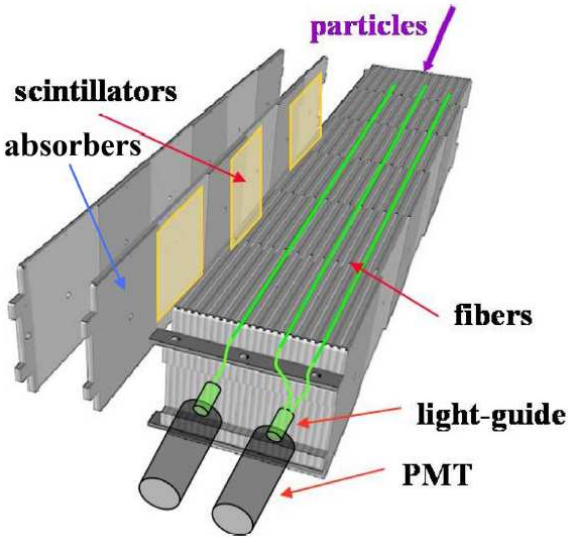


Figure 2.23: The layout of an HCAL module. Reproduced from [40].

sensors, the detector modules consist of alternating absorber and scintillator layers each 1 cm thick, though iron is used as the absorber material. A further difference is that the scintillator and absorber plates are oriented parallel to the beam-line, as shown in figure 2.23. In the longitudinal direction each iron plate is of length λ_i, the interaction length of hadrons in steel. Again, WLS fibres are used to read out the light from the scintillator pads to the PMTs. The granularity of the HCAL modules is varied with distance from the beam-line, as shown in figure 2.22b, by grouping together the WLS fibres from differing numbers of cells to be read by a single PMT. The energy resolution thus achieved is

$$\frac{\sigma_E}{E} = \frac{\sim 69\%}{\sqrt{E}} \oplus \sim 9\%, \tag{2.5}$$

where E is in GeV.

2.2.6 The Muon System

The LHCb Muon tracking system [40, 66] is designed specifically to identify any muons in an event. It consists of five stations, M1-M5, with M1 positioned before the calorimeters, so as to provide more accurate transverse momentum (p_T) measurements to the trigger (discussed in section 2.2.7), and M2-M5 after the calorimeters, as shown in figure 2.3. Iron absorbers 80 cm thick are placed between stations M2-M5, so only high momentum ($p >\sim 6\,\mathrm{GeV}$) muons penetrate to M5. M1-M3 have relatively fine granularity in the horizontal direction, so as to provide accurate p_T measurements, while M4 and M5 are much coarser, and serve mainly to identify high momentum muons. The stations increase in surface area as their

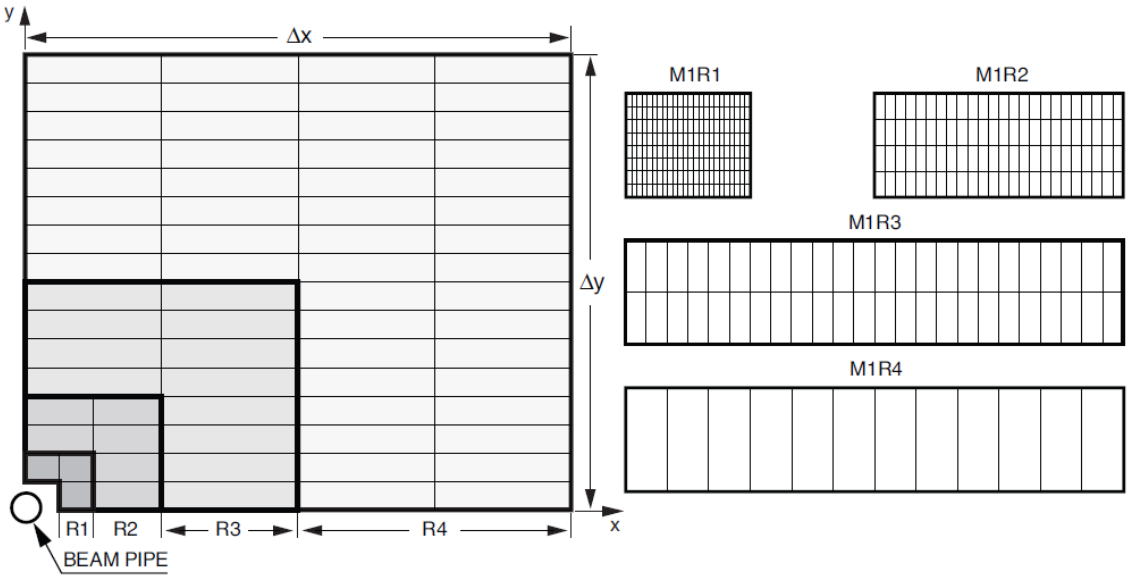


Figure 2.24: The segmentation of a quadrant of a muon station (left) - each rectangle represents one chamber - and the segmentation of individual chambers in the different regions of M1 (right) - each rectangle represents one sensor pad. Reproduced from [40].

distance from the interaction point increases, so as to maintain their angular coverage of 20 mrad to 306 mrad in the horizontal axis, and 16 mrad to 258 mrad in the vertical.

The sensors in M2-M5 are Multi-wire proportional chambers (MWPCs); M1 requires greater radiation hardness due to its proximity to the interaction point, and so uses Gas Electron Multipliers (GEMs) for its innermost region, and MWPCs elsewhere. As with the other subdetectors in LHCb, the granularity of the muon stations varies with distance from the beam-line, to account for the wide variation in particle flux. They are divided into four regions, R1-R4, at increasing distance from the beam-line, as shown in figure 2.24. A sensor in R1 of M1 is $\sim 10 \times 25\,\mathrm{mm}$. The granularity of each region scales with the ratio 1:2:4:8, so as to give each region roughly the same occupancies.

2.2.6.1 Design of the Gas Electron Multipliers

The innermost region, R1, of M1 consists of 12 GEM chambers. The sensitive regions of each GEM chamber comprises two triple-GEM sensors. A triple-GEM sensor consists of an anode and cathode layer, between which lie three GEM foils surrounded by inert gas. A bias voltage is applied to the cathode and three GEM foils. Thus, ionisation electrons produced by a traversing muon in the drift gap between the cathode and GEM foils are accelerated through the GEM foils, and are multiplied by each transition, until they reach the anode and are read out as a signal. A schematic of a triple-GEM sensor is shown in figure 2.25. The

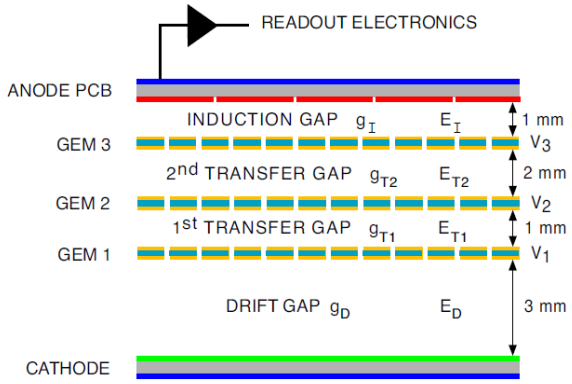


Figure 2.25: A schematic of a triple-GEM sensor, as used in R1 of M1. Reproduced from [40].

gas mixture used is Ar, $\mathrm{CO_2}$ and $\mathrm{CF_4}$ in the ratio 45:15:40, which has been found to give drift times of less than 3 ns.

2.2.6.2 Design of the Multi-Wire Proportional Chambers

The remainder of the muon stations are composed of 1368 MWPCs. Each chamber consists of two cathode plates, spaced 5 mm apart and held at 2600 V; between these plates 30 $\mu\mathrm{m}$ thick gold-plated Tungsten wires are fixed, with 2 mm spacing, in a mixture of Ar, $\mathrm{CO_2}$ and $\mathrm{CF_4}$ in the ratio 40:55:5. Charged particles traversing the chamber create ionisation electrons which are accelerated towards the wires, where they can be read out. The chambers in M1 consist of two layers of MWPCs, while in M2-M5 four layers are used. Figure 2.26 shows a schematic of a four layer MWPC.

2.2.6.3 Performance of the Muon System

The layers in each muon chamber are taken as a logical OR to determine the presence of a muon. In doing so the GEMs achieve an efficiency of more than 96%, while the MWPCs achieve an efficiency of more than 95%. Figure 2.27a shows the efficiency with which μ are correctly identified, as a function of p, while figure 2.27b shows the efficiency with which a π is mis-identified as a μ. This is calculated using pairs of μ from $\mathrm{J/\psi}$ decays, as the $\mathrm{J/\psi}$ can be positively identified by its invariant mass, without using the information from the muon system. The correct ID rate is > 95 % and the mis-ID rate is just a few % for all momenta, demonstrating the excellent performance of the muon system.

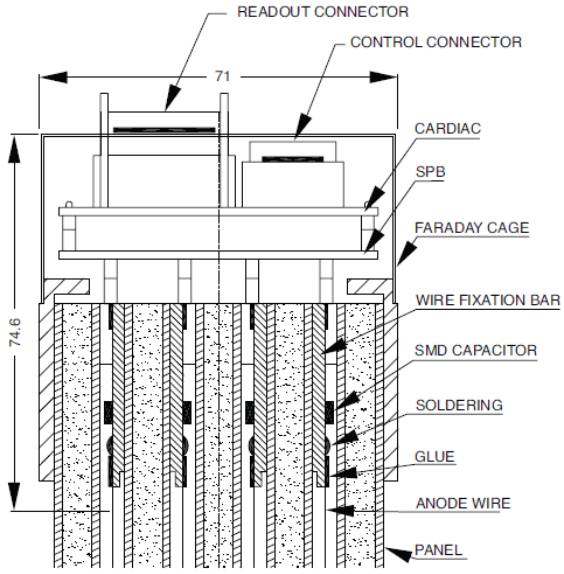
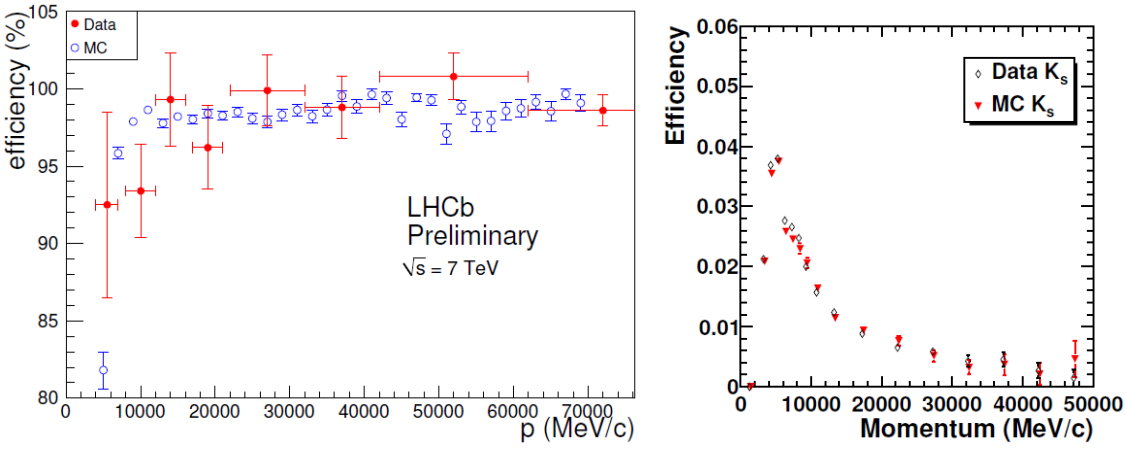


Figure 2.26: A schematic of a four layer MWPC, reproduced from [40].



(a) The efficiency of correctly identifying a μ vs p.

(b) The efficiency of mis-identification of π as μ vs p.

Figure 2.27: PID efficiencies achieved by the muon tracking stations as a function of p. Reproduced from [67].

2.2.7 The Trigger System

Due to the high collision rate provided by the LHC only a fraction of events reconstructed in LHCb can be retained. The decision of whether to keep or discard any given event is made by the Trigger system [40]. The rate of events detected by LHCb during 2011 was $\sim$10 MHz. The Trigger is designed to reduce this to a manageable storage rate of $\sim$3 kHz, by selecting only those events that are most interesting for physics analyses. The limiting factor in the data retention rate is the computing resources available: the data can only be written to permanent storage at a certain rate, and a limited volume of raw storage space is available. Further, the data must be copied to several locations globally, both for backup and to spread the demands on any one storage element, and so must be of a manageable size. Finally, the full offline reconstruction of the data is very CPU intensive, taking $\sim$1.5 s per event. Full datasets must be reconstructed within reasonable time limits, so time cannot be spent reconstructing events that are of limited physics interest.

In order to reduce the retention rate while maximising the signal content of the data the Trigger is designed in three levels: the level-0 (L0) trigger, and the High Level Triggers HLT1 and HLT2. These are operated in a logical AND mode, such that only events passing L0 are processed by HLT1, and only events passing HLT1 are processed by HLT2. Events failing any of these stages are discarded, while those that pass all three are sent to permanent storage.

2.2.7.1 The Level-0 Trigger

The L0 trigger is hardware based and aims to reduce the event rate from $\sim$10 MHz to $\sim$1 MHz, the rate at which data from all detector components can be read out. It is implemented in custom electronics, called the L0 Decision Unit (L0DU), as it is required to make a decision within 1 $\mu\mathrm{s}$. Consequently it only uses information from the calorimeters, muon stations and the VETO stations of the VELO. Due to the relatively high mass of mesons and baryons containing b and c quarks their decay products tend to have high transverse momentum, p_T, and energy, E_T, of several GeV/c. Thus the L0DU uses the calorimeters to look for high E_T clusters, produced by electrons, photons or hadrons, and the muon stations to find pairs of high p_T muons. Events with very high track multiplicity contain too many potential backgrounds to be worth keeping, thus the L0DU also uses the SPD and the VELO VETO stations to perform a quick estimate of how many tracks can be reconstructed by LHCb, and discards events in which this number is too high.

Should an event contain a sufficiently high E_T γ, e, or hadron, or a sufficiently high p_T pair of muons, and sufficiently few reconstructable tracks, it is passed to HLT1; otherwise it

is discarded.

2.2.7.2 The High Level Trigger

The data from all detector elements are read out for events passing L0. These data are then passed, at a rate of $\sim$1 MHz, to the Event Filter computer Farm (EFF), which runs the HLT algorithms. The HLT is implemented in C++ in the software package MOORE, and uses simple, fast reconstruction algorithms. During early data taking in 2010 there were 4400 computer cores in the EFF, and another 4800 were added that October to enable data taking at higher $\mathcal{L}$. In 2011 the $\mathcal{L}$ increased further and 6000 more cores were added, bringing the total to 15,200.

HLT1 attempts to confirm the L0 decision by matching the clusters in the calorimeters or muon stations to tracks in the VELO and tracking stations, or confirm the absence of tracks for clusters made by γ and neutral hadrons. For charged particles the matched track must be confirmed to have sufficiently high p_T, and have an impact parameter greater than 0.11 mm with respect to all primary vertices in the event. HLT1 reduces the retention rate to $\sim$30 kHz, and passes the events selected to HLT2.

HLT2 uses the full detector information to reconstruct and select candidates for the decay channels of interest for physics analyses. It consists of many selection algorithms of many different types, each looking for a specific decay channel and applying different requirements to the candidates found. This is done by the software package DAVINCI [68]. The candidates are made by combining sets of tracks under the hypothesis that they are the decay products ('daughters') of a single 'mother' particle. Their momenta are summed to give the momentum of the mother particle, and a vertexing algorithm run to find the most likely decay point of the mother.

Most selections require that the tracks make a good vertex, *i.e.* pass within a minimum distance of each other; that the sum of the tracks' momenta points sufficiently close to a primary vertex; and that the invariant mass of the mother thus made is within a certain range of the known mass of the particle they aim to select. The full reconstruction performed in HLT2 also allows the selections to use error information on values like flight distance and impact parameters, and thus cut on their χ^2 rather than their raw values, as is done in HLT1. Some selections require the mother particle to be fully reconstructed, *e.g.* $\mathrm{D}^0 \to \mathrm{K}\pi$, and are thus called 'exclusive' selections; others, known as 'inclusive' selections, only require the mother to be partially reconstructed, *e.g.* $\mathrm{B}^0 \to \mathrm{J}/\psi\, X$, where X can be any particle, and need not necessarily be reconstructed.

Each selection in HLT2 is taken in a logical OR to decide whether to keep or discard an event, thus reducing the event retention rate to $\sim$3 kHz. Events passing HLT2 are sent to

permanent storage, to be fully reconstructed offline and used in physics analyses.

2.2.8 Offline Data Processing and Simulation Software

2.2.8.1 Offline Data Processing

Once the raw data selected by the trigger are stored, they need to be reconstructed fully before any physics analyses can be performed. The offline reconstruction is performed by the software package BRUNEL. This takes the raw hits and clusters read out from the detector and produces tracks and primary vertices, as was described for the VELO in section 2.2.1.2. It then associates these tracks with Cherenkov rings in the RICH detectors, as described in section 2.2.2.2, and with clusters in the calorimeters and muon stations, and produces a likelihood for each track being π, K, e, μ or p. The track and vertex fitting, and the PID algorithms used by BRUNEL are very rigorous, and result in a significant improvement in the resolutions of measurable parameters over the fast algorithms used in MOORE. However, this makes reconstruction process very CPU intensive, taking $\sim$1.5 s per event. Thus all the data from stable running periods are reconstructed in single sessions at times when stable versions of the reconstruction algorithms and detector alignment information are produced. The processing session for data from the 2011 run took several months to complete. A 'reprocessing' involves running the reconstruction on data that have been previously reconstructed, and is only performed when it is deemed entirely necessary, *e.g.* when a new detector alignment has been produced that significantly improves the data quality. The files output by BRUNEL contain all the information concerning the fully reconstructed PVs, tracks and their PID likelihoods, and are saved to permanent storage.

With the fully reconstructed tracks with full PID information available, searches can be performed for the particles of interest, such as B and D mesons. This is performed by the software package DAVINCI, as was described for HLT2 in section 2.2.7.2: sets of tracks are combined under the hypothesis that they were made by the daughters of a single mother particle, and the mother particle's momentum and trajectory defined as the sum of those of the daughters. Similarly to HLT2 a set of selection algorithms are run on the data, each looking for a separate decay channel and applying different selection criteria to the candidates found. These are known as 'stripping' selections, and differ from HLT2 in that they are performed on the fully offline reconstructed data, and can exploit the significantly better resolution on measurable parameters. The stripping selections serve to confirm the HLT2 decisions, or apply tighter selections to remove backgrounds more effectively, using all the information available offline. The stripping algorithms are run directly after the data have been reconstructed, during a data processing or reprocessing session. The stripping algo-

rithms are grouped broadly by the specific decays they analyse, *e.g.* B meson decays, or D meson decays. Each group produces its own output files which constitute a much reduced dataset containing only the candidates that are of interest to the specific analyses for which the stripping selections in that group were designed. DAVINCI can then be run again by the end user to extract the parameters of interest for their analyses from the candidates thus stored, *e.g.* the mass or proper decay time of the mother particle, so that fits can be performed to extract the underlying physics parameters.

2.2.8.2 Simulation Software

Although the analyses presented in this thesis use data driven techniques for the measurement methodology and to evaluate systematic uncertainties, accurate simulation of the whole collision and detection process is important to verify many physics analyses performed at LHCb. For this reason a full simulation of the LHCb detector has been written using the software package GEANT4 [69]; this is contained in the software package GAUSS. GAUSS uses the physics simulation packages PYTHIA [70] and EVTGEN [71] to simulate proton-proton collisions like those provided by the LHC. The particles thus produced are then propagated through the simulated detector and their interactions with the detector material, *e.g.* scattering and deposition of energy, is also simulated. The response of the detector elements to these energy deposits is simulated by the package BOOLE. The energy is transformed into signals in the detector sensors, and any pre-processing performed by the detector readout electronics is also simulated, resulting in raw data in the same format as is read out from the detector in reality.

From there, the simulated data can be treated identically to data read out from the detector: the trigger can be run using MOORE; the data can be fully reconstructed using BRUNEL, and parameters of interest extracted by DAVINCI. The added benefit of the simulated data is that the particles generated by PYTHIA can also be saved. By accessing this information and comparing it to the hits, tracks and particles reconstructed in the detector one can easily assess reconstruction and selection efficiencies, *etc.*; in addition one can confirm the validity of any physics analyses by comparing the values of the physics parameters extracted from the reconstructed data to those used in generating the collisions and their products. Thus such simulation is essential to verify the performance of the detector and the analyses performed using the data it records.

2.3 Summary

This chapter has described in detail the design and performance of the LHCb detector, one of the experiments on the LHC at CERN. Each of the subdetectors comprising LHCb was discussed in turn and details given as to their construction. Where possible, an evaluation of their performance using data collected so far by LHCb was shown, demonstrating the excellent performance of the detector as a whole. The trigger system, used to select which events to keep and which to discard, was described. The data processing required to perform physics analyses, and the software packages that perform it, was also discussed. Finally, a brief description was given of the software packages used in performing full simulations of the physics processes at the LHC and the response of the LHCb detector. How these are used to verify the performance of the detector and the accuracy of the physics analyses performed at LHCb was also detailed.

Chapter 3

Measurement and Characterisation of Impact Parameter Resolutions

The impact parameter (IP) of a track with respect to a certain space point (such as a primary vertex (PV)) is defined as the distance of closest approach of the track to the point, as is shown in figure 3.1a. Tracks made by daughters of long lived particles, that are produced at a displaced decay vertex, consequently tend to have much larger IPs with respect to any PVs in an event, as is shown in figure 3.1b. Further the IP of the daughter particle is strongly correlated to the proper decay time of the mother particle:

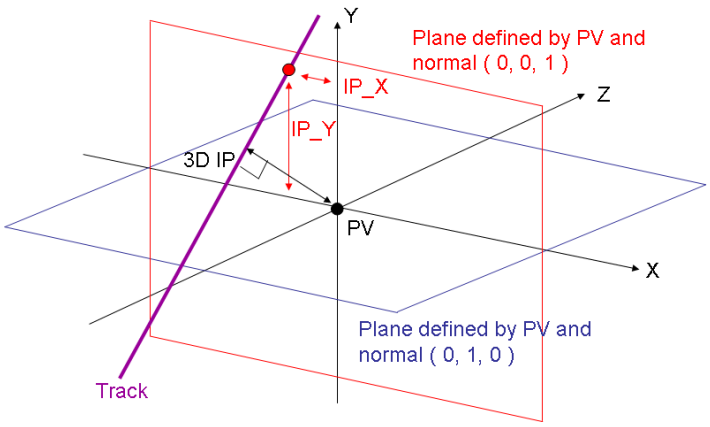
$$\begin{aligned} \text{IP}_{3D} &= |\vec{IP}| \\ &= |\underline{DV} - \underline{PV}| \sin\theta \\ &= \frac{|\vec{p}_{mother}|\tau}{m_{mother}} \sin\theta, \end{aligned} \tag{3.1}$$

where $\underline{PV}$ is the position of the PV, $\underline{DV}$, $\vec{p}_{mother}$, m_{mother} and τ are the decay vertex, momentum, mass and proper decay time of the mother particle. θ is the opening angle of the daughter particle with respect to the mother particle, defined as

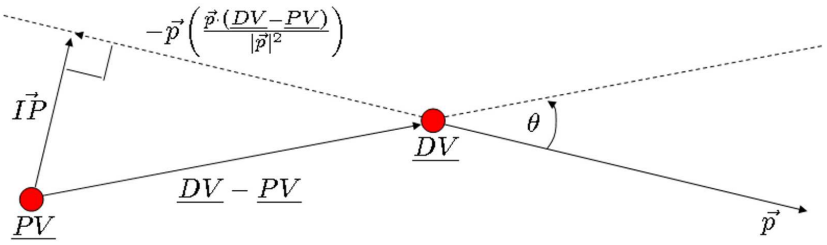
$$\cos\theta = \frac{\vec{p}_{daughter} \cdot \vec{p}_{mother}}{|\vec{p}_{daughter}||\vec{p}_{mother}|}, \tag{3.2}$$

where $\vec{p}_{daughter}$ is the momentum of the daughter particle.

When looking for a specific species of particle all possible combinations of tracks are tried to see if any of them fulfil the criteria for having been made by the daughters of a particle of that species, as described in section 2.2.8. Thus, one type of background that any analysis like those conducted at LHCb must deal with is from combinations of random tracks made by particles that don't have the same mother particle, but happen to produce a candidate that looks like signal. Such candidates are known as 'combinatorial backgrounds'.



(a) The definition of the impact parameter of a track with respect to a primary vertex.



(b) The dependence of impact parameters of daughter particles on the flight distance of their mother particle. Here $\underline{DV}$ is the decay vertex of the mother particle, and $\vec{p}$ the momentum of the daughter particle.

Figure 3.1

Selection criteria are applied to all candidates to minimise the background contamination. The majority of tracks detected by LHCb are made by direct products of the fragmentation of the colliding protons, which are produced exactly at the PV. Such tracks will thus have small IPs with respect to the PV. The fact that the daughters of long lived particles tend to have very large IPs thus makes the daughter IP a very powerful selection criterion when trying to exclude backgrounds. Cutting on χ^2_{IP} is even more powerful. The χ^2_{IP} of a track is defined as the increase PV χ^2, as calculated using equation 2.2, when the track is added to the PV fit (note that this is not exactly equal to IP^2/σ^2_{IP} as the position of the PV changes with the addition of the track). Tracks made by particles produced at the PV have χ^2_{IP} consistent with having IP $= 0$, while those made by particles not produced at the PV tend to have much larger χ^2_{IP}.

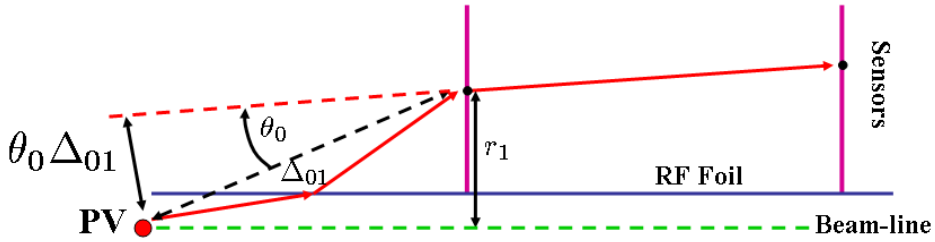
As χ^2_{IP} cuts are so common it is very important to verify that the uncertainties calculated for IP measurements are accurate. To do this one must have detailed measurements of IP resolutions, and a full understanding of what affects the resolution on such measurements. This chapter presents a method to evaluate the resolution with which IP measurements can be made at LHCb, without the use of simulation, and the results of its use. Detailed comparisons with results obtained from simulated data are then made. Section 3.1 discusses the effects that contribute to IP resolutions; section 3.2 discusses the method used to evaluate IP resolutions; section 3.3 demonstrates how the expected behaviour is observed, and makes a comparison between real and simulated data; and finally, section 3.6 presents a more detailed analysis and comparison, including checks of the error parametrisation.

3.1 Contributing Factors

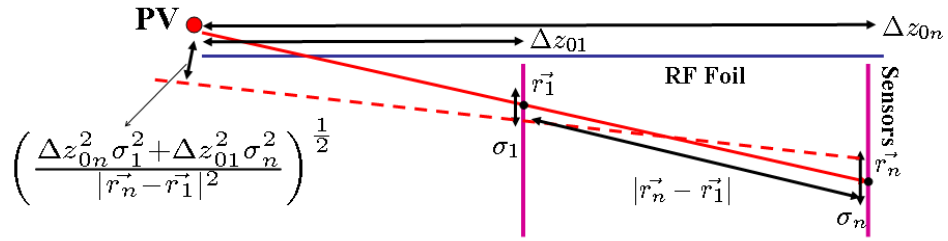
There are three main factors that cause a reconstructed track to deviate from the original trajectory of the particle that made it, and thus influence the resolution with which IP measurements are made: the resolution with which the position of single hits on the track can be determined (the 'detector resolution' or 'single hit resolution'); the amount of detector material through which particles must pass; and the distance between the first hit on a reconstructed track and the interaction point [72]. The effects of these factors are shown in figure 3.2.

A particle passing through the detector material interacts with the electrons and nucleons therein and is deflected through a small angle θ_0. The distribution of the scattering angles is assumed to be Gaussian with mean zero and σ given by [73]

$$\sigma_\theta = \frac{0.0136}{p}\sqrt{x/X_0}[1 + 0.038\ln(x/X_0)], \tag{3.3}$$



(a) A diagram of the VELO showing the effects of multiple scattering in detector material and extrapolation distance to the interaction point on impact parameter resolution. The solid red lines show the true trajectory of the particle, while the dashed red lines show the reconstructed trajectory of the particle. When a particle passes through the detector material its trajectory is displaced by a small angle θ_0. When the trajectory is then extrapolated back to the PV over the distance Δ_{01} this causes a displacement of the IP of size $\theta_0\Delta_{01}$.



(b) A diagram of the VELO showing the effects of single hit resolution and extrapolation distance to the interaction point on impact parameter resolution. The solid red lines show the true trajectory of the particle, while the dashed red lines show the reconstructed trajectory of the particle. As the reconstructed hit positions are slightly displaced from the true point of intersection of the particle with the sensors this causes a displacement to the IP when the trajectory is extrapolated to the PV. The size of this displacement is dependent on the single hit resolutions on each hit, the extrapolation distance to the PV, and the distance between the two hits.

Figure 3.2

where p is the momentum of the particle in GeV, x is the distance travelled by the particle in the material, and X_0 is the radiation length of the material. The resulting contribution to the IP resolution is the scattering angle multiplied by the distance between the interaction point and the first hit on the reconstructed track, Δ_{01}.

To evaluate the effects of single hit resolution one must consider how a track is reconstructed from the hits in the detector. If a track is reconstructed using linear interpolation between two space points $\vec{r_1}$ and $\vec{r_2}$ its trajectory is given by

$$\vec{r}(t) = \vec{r_1} + (\vec{r_2} - \vec{r_1})t, \tag{3.4}$$

where t is the parameter of the line. Assuming the PV is at the origin the value of t at the point of closest approach of the track to the PV (t_{POCA}) is found by minimising $|\vec{r}(t)|^2$ with respect to t, which gives

$$t_{POCA} = \frac{-\vec{r_1} \cdot (\vec{r_2} - \vec{r_1})}{|\vec{r_2} - \vec{r_1}|^2}. \tag{3.5}$$

Thus the impact parameter is

$$\begin{aligned} \vec{IP} &= \vec{r}(t_{POCA}) \\ &= \vec{r_1} + (\vec{r_2} - \vec{r_1})t_{POCA} \\ &= \vec{r_1} + (\vec{r_2} - \vec{r_1}) \left(\frac{-\vec{r_1} \cdot (\vec{r_2} - \vec{r_1})}{|\vec{r_2} - \vec{r_1}|^2} \right) \\ &= \frac{1}{|\vec{r_2} - \vec{r_1}|^2} \left(\vec{r_1} |\vec{r_2} - \vec{r_1}|^2 - (\vec{r_2} - \vec{r_1}) (\vec{r_1} \cdot (\vec{r_2} - \vec{r_1})) \right). \end{aligned} \tag{3.6}$$

Adopting a 2D r-z coordinate system such that $\vec{r_n} = \begin{pmatrix} r_n \\ z_n \end{pmatrix}$ this simplifies to

$$\vec{IP} = \frac{r_1(z_2 - z_1) - z_1(r_2 - r_1)}{|\vec{r_2} - \vec{r_1}|^2} \begin{pmatrix} z_2 - z_1 \\ -(r_2 - r_1) \end{pmatrix}. \tag{3.7}$$

Requiring that the track passes through PV (the origin) we have that $\vec{r_2} = \frac{z_2}{z_1}\vec{r_1}$, so $r_2 = \frac{z_2}{z_1} r_1$, which gives $\vec{IP} = \vec{0}$.

The finite single hit resolution results in the reconstructed hit positions being slightly displaced from the true point of intersection of a particle with the sensors. If one neglects the z resolution but applies an r resolution such that $r_n \to r_n + \delta r_n$, requiring that the track passes through the PV in the absence of the resolution effects gives

$$\begin{aligned} \vec{IP} &= \frac{(r_1 + \delta r_1)(z_2 - z_1) - z_1(\frac{z_2}{z_1} r_1 + \delta r_2 - r_1 - \delta r_1)}{|\vec{r_2} - \vec{r_1}|^2} \begin{pmatrix} z_2 - z_1 \\ -(\frac{z_2}{z_1} r_1 + \delta r_2 - r_1 - \delta r_1) \end{pmatrix} \\ &= \frac{\delta r_1 z_2 - \delta r_2 z_1}{|\vec{r_2} - \vec{r_1}|^2} \begin{pmatrix} z_2 - z_1 \\ -(\frac{z_2}{z_1} r_1 + \delta r_2 - r_1 - \delta r_1) \end{pmatrix}. \end{aligned} \tag{3.8}$$

Thus

$$\begin{aligned}\left|\vec{IP}\right| &= \left(\frac{\left(\delta r_1 z_2 - \delta r_2 z_1\right)^2}{\left|\vec{r_2} - \vec{r_1}\right|^4}\left((z_2 - z_1)^2 + (\frac{z_2}{z_1}r_1 + \delta r_2 - r_1 - \delta r_1)^2\right)\right)^{\frac{1}{2}} \\ &= \frac{\delta r_1 z_2 - \delta r_2 z_1}{\left|\vec{r_2} - \vec{r_1}\right|}. \end{aligned} \tag{3.9}$$

Figure 2.9a shows that the single hit residuals are well described by a single Gaussian, thus the resolution parameters δr_n follow Gaussian distributions with mean zero. When Gaussian random variables are added or subtracted the resulting variable is Gaussian with σ^2 equal to the the sum of the variances of the underlying variables, and mean equal to the sum of their means. Thus the contribution to the IP resolution that results from the single hit resolution will be Gaussian with mean zero and σ given by

$$\sigma^2_{hit} = \frac{\sigma_1^2 z_2^2 + \sigma_2^2 z_1^2}{\left|\vec{r_2} - \vec{r_1}\right|^2}, \tag{3.10}$$

where σ_n is the r hit resolution at $\vec{r_n}$. Figure 2.9 shows that the r hit resolution is linearly dependent on the strip pitch. One can crudely approximate the variation in strip pitch of the VELO sensors as being proportional to the radial distance from the beam-line. Under this approximation one has that $\sigma_2 = \frac{r_2}{r_1}\sigma_1 = \frac{z_2}{z_1}\sigma_1$, and so

$$\begin{aligned}\sigma^2_{hit} &= \frac{2\sigma_1^2 z_2^2}{\left|\vec{r_2} - \vec{r_1}\right|^2} \\ &= f_{12}^2 \sigma_1^2, \end{aligned} \tag{3.11}$$

where $f_{12} \equiv \sqrt{2}\frac{z_2}{|\vec{r_2} - \vec{r_1}|}$ is the 'extrapolation factor' [72]. As the PVs reconstructed by LHCb are not exactly at the origin one must replace z_2 with $z_2 - z_{PV} \equiv \Delta z_{02}$, and so $f_{12} = \sqrt{2}\frac{\Delta z_{02}}{|\vec{r_2} - \vec{r_1}|}$.

Of course, tracks in the VELO are reconstructed from hits in more than two VELO stations. Thus a point in the nth station downstream of the first hit, $\vec{p_n}$, is used instead of $\vec{r_2}$ to calculate the extrapolation factor f_{1n}. Using that $|\vec{r_n} - \vec{r_1}| \overset{n\to\infty}{\to} \Delta z_{02}$, one has that $f_{1n} \overset{n\to\infty}{\to} \sqrt{2}$, and so the contribution of hit resolution to IP resolution is reduced as n increases. Particles are also scattered as they pass through the sensors downstream of the first hit on their corresponding track. Thus, exactly what n to use is dependent on the p of the particle. As each VELO sensor is 300 μm thick and the radiation length of silicon is 93.7 mm a VELO module comprises 0.64 % of one radiation length. In [72] it is reasoned that, as the VELO stations are spaced 30 mm apart about the interaction point, equation 3.3 finds that the displacement due to multiple scattering of a particle as it travels between two VELO stations is $\frac{0.032}{p}$ mm, for p in GeV. As the mean single hit resolution is $\sim$8 μm

this displacement is smaller than the hit resolution for $p > 4\,\mathrm{GeV}$. Thus one should use $n = 2 + \lfloor p/4 \rfloor$, for p in GeV, to calculate f_{1n}.

In 2D geometry the IP resolution can thus be described by a Gaussian with mean zero and σ given by

$$\begin{aligned}\sigma_{IP}^2 &= \sigma_{hit}^2 + \sigma_\theta^2 \Delta_{01}^2 \\ &= f_{1n}^2 \sigma_1^2 + \left(\frac{0.0136}{p} \sqrt{x/X_0} [1 + 0.038 \ln(x/X_0)] \right)^2 \Delta_{01}^2 \\ &= f_{1n}^2 \sigma_1^2 + \left(\frac{0.0136}{p_T} \sqrt{x/X_0} [1 + 0.038 \ln(x/X_0)] \right)^2 r_1^2, \end{aligned} \tag{3.12}$$

using $\Delta_{01}/p = r_1/p_T$, where r_1 is the radius of the first hit on the track.

In 3D geometry an IP has two degrees of freedom - three as it is a distance in 3D space, minus one from the requirement of being taken at the point of closest approach to the PV. The two underlying variables have identical Gaussian distributions with σ given by equation 3.12, and so the measured IP resolution is decoupled into its 1D x and y components, as shown in figure 3.1a. Due to the forward geometry of LHCb the z component is negligible. An IP measurement in 3D space is thus simply the sum in quadrature of its x and y components, $\sqrt{\mathrm{IP}_x^2 + \mathrm{IP}_y^2}$. The mean offset of such a measurement from its true value is given by the resolution on the 1D components multiplied by $\sqrt{\pi/2}$.

Measuring the σ of the distributions of IP_x and IP_y as a function of $1/p_T$ one thus expects a roughly linear distribution with y-intercept dependent on the single hit resolution, and gradient dependent on the detector's material budget. f_{1n} has some dependence on momentum due to the dependence of n on momentum. However, as this is a much weaker dependence than for multiple scattering it contributes little to the gradient of the distribution.

3.2 Measurement Methodology

As stated previously, the vast majority of particles detected by LHCb are produced exactly at the point of the proton-proton collision, and thus only have non-zero IP due to resolution effects. Assuming the fraction of particles produced at displaced decay vertices and the fraction of mis-reconstructed tracks to be negligible, one can examine IP resolutions simply by measuring the IP of all reconstructed tracks with respect to the PVs in an event.

As there is a non-zero resolution on the position of reconstructed PVs this also makes a small contribution to the measured IP. The position of a PV is calculated by minimising the χ_{IP}^2 of a set of tracks with respect to the PV, as described in section 2.2.1.2 and in [51]. Consequently each track included in the fit has some influence on the position of the PV. To correctly account for this each track must be excluded from the PV fit and the PV refitted,

so that the track has no influence on the PV position, before the IP of the track is calculated. The contribution of PV resolution to the measured IP can be minimised by using only PVs which have been fitted using a large number of tracks, and thus have much smaller resolution on their position. Alternatively, one can use an independent measurement of the PV position resolution to decouple it from the measured IP.

In order to minimise the contribution from mis-reconstructed 'ghost' tracks and poorly reconstructed tracks that are excluded from most analyses, quality requirements are placed on the tracks used. These are

$$\begin{aligned} \chi^2(\text{track})/NDOF(\text{track}) &< 4 \\ \text{N. VELO R hits} &> 5 \\ \text{N. TT hits} &> 0 \\ p_T &> 300\,\text{MeV} \\ p &< 500\,\text{GeV}, \end{aligned} \tag{3.13}$$

where 'N. VELO R hits' is the number of hits on the track in VELO R sensors, and'N. TT hits' is the number of hits on the track in the 'Tracker Turicensis' (the first of the tracking stations downstream of the VELO). The $\chi^2/NDOF$ cut simply requires the tracks to be well fitted, and similarly for the minimum requirement on the number of VELO hits used in the track. An analogous cut on the minimum number of hits is used in HLT1. The maximum limit on p excludes any tracks that are very close to being straight lines, and thus have very poor momentum resolution, while the minimum limit on p_T removes tracks that are excessively curved, and thus likely to be mis-reconstructed. Finally, requiring at least 1 hit on the track in the TT reduces the likelihood that the VELO segment of the track is mis-reconstructed. Further, only tracks that have been reconstructed in both the VELO and the tracking stations downstream of the dipole magnet are used, as these have the most reliable momentum measurements. On simulated data these reduce the fraction of ghost tracks from $\sim$13 % to $\sim$1.7 %, and the fraction of tracks from particles produced at displaced vertices from $\sim$9.2 % to $\sim$3.9 %.

Having done this, one can then examine the distributions of IP_x and IP_y and perform Gaussian fits to extract the resolution.

3.3 Basic Characterisation

In this section the effects of PV resolution are suppressed by requiring the PV to be reconstructed with more than 25 tracks. This is a reasonably soft cut, as most PVs are reconstructed with 50 or more tracks, but ensures that the PV resolution is generally much smaller

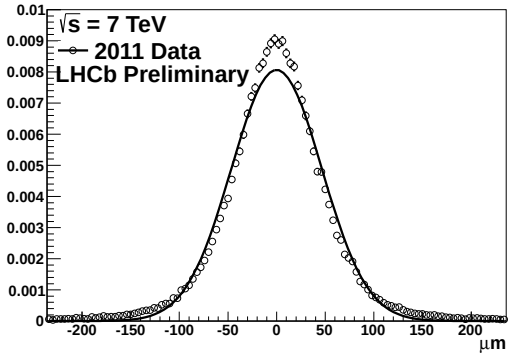


Figure 3.3: The distribution of IP_x residuals for tracks with $1.35 < 1/p_T[\,\mathrm{GeV}/c] < 1.5$ with a single Gaussian fit. Made using 2011 data.

than the IP resolution. To avoid any mis-association of tracks with PVs only events with one reconstructed PV are used. Figure 3.3 shows the distribution of IP_x residuals in the highest occupancy bin of $1/p_T$ overlaid with the single Gaussian fit. While the single Gaussian does not describe the data exactly it is sufficiently good as to give an estimate of the resolution. Figure 3.4 shows the σ of Gaussian fits to the distributions of IP_x and IP_y in bins of $1/p_T$, with data taken in early 2011 shown in black. The resolutions of IP_x and IP_y follow almost identical distributions and a strong, roughly linear, dependence on $1/p_T$ is seen, as predicted by equation 3.12. The results of linear fits to the distributions are also shown. As is also shown in section 2.2.1.3 for tracks with $p_T > 1\,\mathrm{GeV}$ the IP resolutions are $\leq \sim 35\ \mu\mathrm{m}$, showing the excellent performance of the VELO in this respect. The same measurement on simulated data is shown in red. The differences between IP resolutions for real and simulated data are discussed in sections 3.4 and 3.6.

Due to the complicated shape of the RF-foil the material budget varies greatly across different regions of the VELO. In particular, there is significantly more material in the region in which the two halves of the VELO overlap, as do the two sides of the RF-foil. Figure 3.5a shows the mean amount of material, in radiation lengths, intersected by tracks between the PV and their first hit, as a function of the azimuthal angle ϕ of the tracks. The material budget is obtained by extrapolating tracks through the simulated description of the detector, for both real and simulated data, and averaging the amount of material intersected by the tracks in each bin. Thus, the value obtained may not be entirely accurate for real data, but at least gives a rough estimate. Also shown are the regions in ϕ that correspond roughly to the two separate halves of the VELO- the A and C sides. In the overlap region, about $\phi = \pm\pi/2$, the tracks intersect $\sim$3.5 times as much material before their first hit as tracks outside the overlap region. Figure 3.5b shows the resolution of IP_x as a function of ϕ, which clearly mirrors the distribution of the material. As tracks with a range of p_T values are used to make this plot

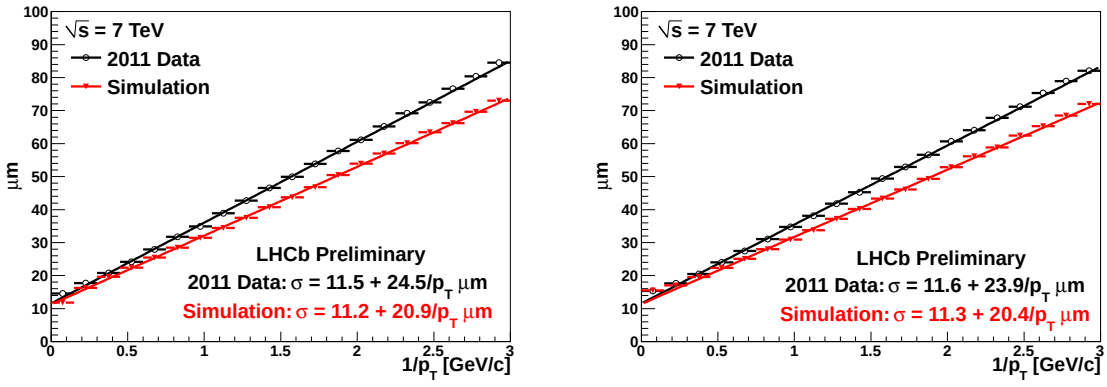
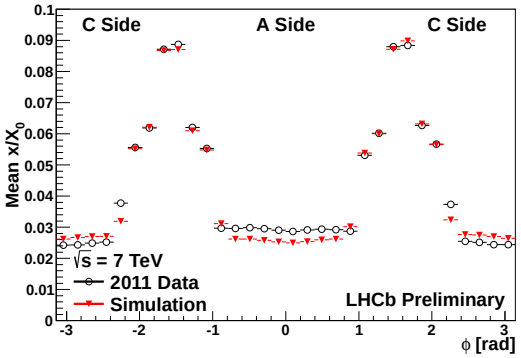


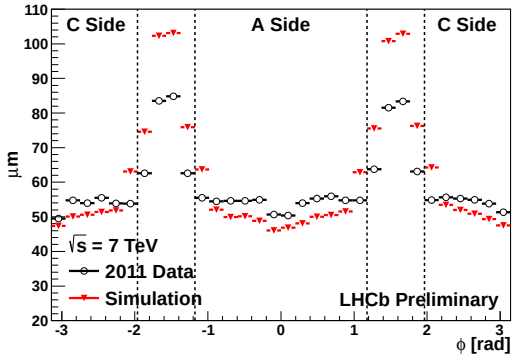
Figure 3.4: The resolution of (a) IP_x and (b) IP_y as a function of $1/p_T$, comparing measurements made using 2011 data to those made using simulated data. This is measured as the σ of a single Gaussian fit to the residual distributions in each bin, an example of which is shown in figure 3.3.

the single Gaussian fits performed to obtain the resolution do not fit the data very well, but provide an estimate of the average resolution. Plotting the resolution as a function of $1/p_T$ for tracks in the overlap region, defined as $\phi \in [\frac{-5\pi}{8}, \frac{-3\pi}{8}] \cup [\frac{3\pi}{8}, \frac{5\pi}{8}]$, one obtains figure 3.5c, while tracks outside the overlap region give figure 3.5d. The y intercepts of the distributions in these two figures are very similar, while the gradient in the overlap region is considerably larger. This is in agreement with the predictions of equation 3.12, as the material budget only affects the gradient of the distribution as a function of $1/p_T$.

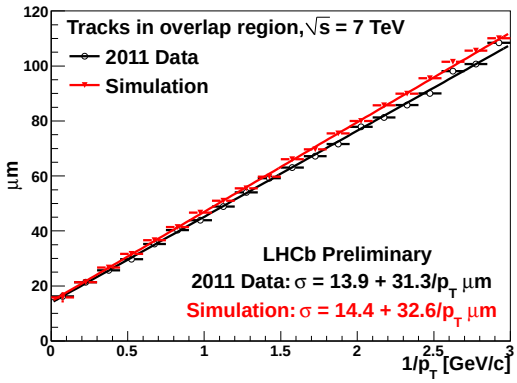
The fast reconstruction algorithms used in the HLT, described in section 2.2.1.2, result in a slightly worse single hit resolution than that of the rigorous methods used offline. Thus, to evaluate the effect of varying the single hit resolution one can compare the IP resolutions attained in the HLT to those found offline, as shown in figure 3.6a. Both the HLT reconstruction methods yield larger y-intercepts than the offline method, as a result of the poorer hit resolution, again in agreement with equation 3.12. The simplest 'line of best fit' method used in the HLT, labelled 'HLT Unfitted' in the figure, does not take into account the scattering of particles as they pass through detector; thus, multiple scattering contributes more strongly to the IP resolution, resulting in a steeper gradient as a function of $1/p_T$. The fast Kalman filter method used in the HLT, labelled 'HLT Fitted', does not suffer from this short-coming. Hence, it agrees well with the offline resolutions at low p_T (high $1/p_T$), and differs only at high p_T due to the poorer hit resolution. It in fact appears slightly better than the offline resolutions at low p_T. This may be due to the fact that the HLT pattern recognition has a lower track finding efficiency than the offline algorithm, particularly at low momentum, and



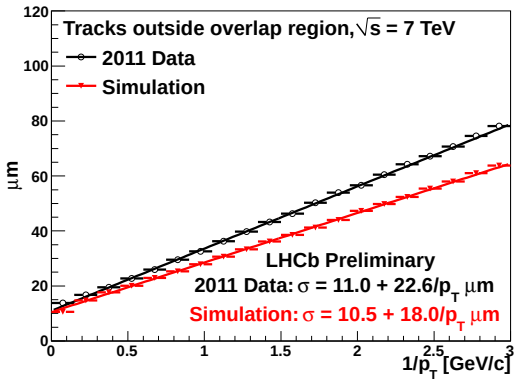
(a) The mean material budget, in radiation lengths, in the simulated description of the detector between the interaction point and the first hit on a track, as a function of the azimuthal angle ϕ. This is calculated by extrapolating tracks from both real and simulated data through the simulated description of the detector, and so does not necessarily represent the material budget of the VELO in reality.



(b) The average resolution of IP_x as a function of the azimuthal angle ϕ of the tracks, integrating over p_T. The dashed lines show the divisions used to select the overlap region and its complement when making figures (c) and (d).

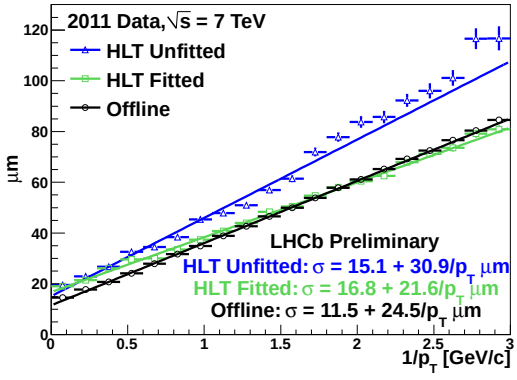


(c) The resolution of IP_x vs $1/p_T$ for tracks in the overlap region, $\phi \in \left[\frac{-5\pi}{8}, \frac{-3\pi}{8}\right] \cup \left[\frac{3\pi}{8}, \frac{5\pi}{8}\right]$.

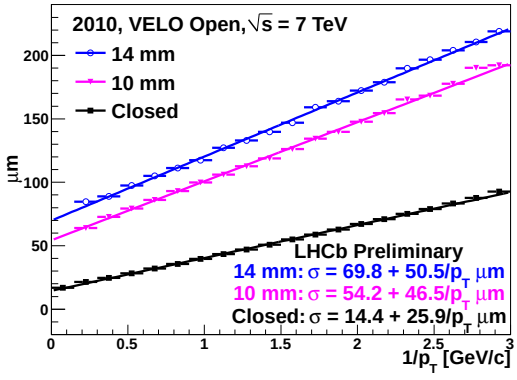


(d) The resolution of IP_x vs $1/p_T$ for tracks outside the overlap region, $\phi \in \left(-\pi, \frac{-5\pi}{8}\right) \cup \left(\frac{-3\pi}{8}, \frac{3\pi}{8}\right) \cup \left(\frac{5\pi}{8}, \pi\right)$.

Figure 3.5



(a) The resolution of IP$_x$ as a function of $1/p_T$ from 2011 data, comparing the HLT reconstruction algorithms to those used offline.



(b) The resolution of IP$_x$ as a function of $1/p_T$ comparing data taken in 2010 with the VELO at different openings.

Figure 3.6

so only finds the better quality tracks.

At the beginning of the 2010 run the stability of the LHC beams was not verified, and so some data was taken with the VELO open. As the beam stability improved the VELO was gradually closed. This affords the opportunity to check the effect of varying extrapolation distance on IP resolutions, by measuring them for each value of the VELO opening, as shown in figure 3.6b. As track and vertex reconstruction is more difficult with the VELO open the minimum number of R hits in the VELO was reduced to 3 and the minimum number of tracks used to reconstruct the PV reduced to 10 to make these plots. The effects of single hit resolution and multiple scattering are both amplified by the extrapolation distance, thus the IP resolution is increased uniformly at larger VELO openings. The sensitive area of the VELO sensors starts at 8.2 mm from the beam-line, thus with the VELO open at 10 mm it starts at $\sqrt{8.2^2 + 10^2} \approx 12.9$ mm, and for 14 mm opening at $\sim$16.2 mm. It is apparent that the gradient of the distribution as a function of $1/p_T$ scales roughly linearly with the distance to the sensitive region, while the y-intercept has a more complicated dependence. This behaviour is also predicted by equation 3.12.

Thus the expected behaviour of IP resolutions with varying single hit resolution, material budget and extrapolation distance has been shown to agree qualitatively with that predicted by equation 3.12. A quantitative comparison is made in section 3.6.

3.4 Comparison to Simulated Data

From figures 3.4 and 3.5 it is clear that IP resolutions behave in the same manner in simulated data as in real data, though there are some striking differences. The y-intercepts of the distributions as a function of $1/p_T$, shown in figure 3.4, are very similar for simulated data and 2011 data, indicating that the hit resolutions agree reasonably well between them. This has been confirmed by independent measurements of the single hit resolution [50]. The large difference in the gradients of these distributions, however, suggests an issue with the detector material in the simulation. Figure 3.5a shows that the tracks from real and simulated data intersect roughly the same amount of material in the simulated description of the detector. The small differences observed are likely due to differences in the distributions of tracks and PVs between real data and simulation. However, figures 3.5c and 3.5d show that there is a good agreement between simulated and 2011 data in the overlap region of the VELO, and a large discrepancy in their gradients as a function of $1/p_T$ elsewhere. This could mean that there is material missing from the simulated description of the detector, that the distribution of the material is incorrect, or that the model of multiple scattering in the detector material is incorrect in the simulation.

To explain this discrepancy, any material missing from the simulated description of the VELO must be outside the overlap region. However, the gradient of the distribution as a function of $1/p_T$ depends on $\sqrt{x/X_0}$. Thus, it would require $\sim$50 % more material outside the overlap region in reality than in the simulation to account for the difference in gradients between real and simulated data. This stimulated detailed studies into the simulation of the VELO [74]. Only small discrepancies in the amount of material were found, though some differences in the shape of the RF-foil were apparent. One error discovered was that the RF-foil was 250 μm thick in the simulation, which was the original design value. This was corrected to the actual manufactured value of 300 μm. This increased the gradient of the IP resolution as a function of $1/p_T$ only slightly. The RF-foil is known to in fact be $\sim$15 % heavier in the simulation than in reality after this correction. This explains why the resolutions on simulated data are actually slightly worse than on real data in the overlap region. Tomography of the VELO has also been performed using real and simulated data, by examining the distributions of vertices formed by interactions in the VELO material [75]. Again, small discrepancies are evident between data and simulation, particularly in the shape of the RF-foil, but it is not clear if this is sufficient to account for the difference in IP resolutions. Studies of the effects of using a more accurate model of the RF-foil in the simulation are currently under way.

The modelling of multiple scattering in GEANT has also been studied in detail [76, 77],

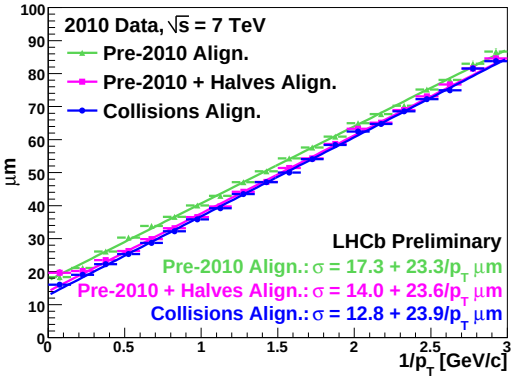
but only very small issues have been found. Thus the source of the discrepancy is still under investigation. More detailed comparisons of IP resolutions on real and simulated data are made in section 3.6, with particular attention paid to the implications for the material budget.

3.5 The Effects of VELO Sensor and Module Alignment

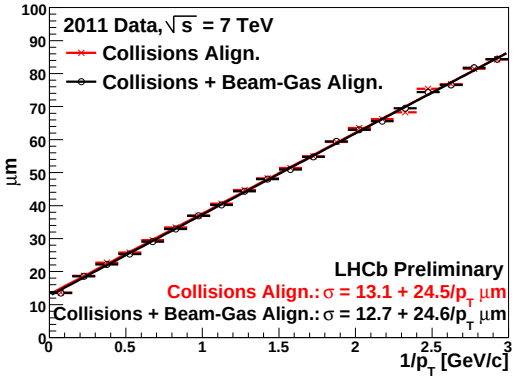
In order to accurately determine the position of hits in the VELO the positions and relative alignment of its component parts must be known very precisely. This requires knowledge of the positions of the two halves of the VELO, the modules within each half, and the sensors in each module. The alignment is intended to be known to a precision below the single hit resolution of the detector, so that the single hit resolution is not degraded by misalignments. Systematic offsets of the assumed component positions from their true positions cause biases in the hit resolutions, displacing the mean residual from zero. The VELO module positions were measured to a precision of $\sim$10 μm after its assembly and before its installation. These measurements were made at room temperature. For operation the VELO sensors are cooled to their operational temperature of $\sim$-5 °C. The baseplate to which the modules are attached is maintained at $\sim$20°C. Consequently the VELO alignment needed to be determined indirectly after its installation and adjustment to operational conditions, to account for any movement of its components. Further, the VELO is closed and opened at the beginning and end of each fill, hence both long- and short-term alignment stability needs to be monitored.

The alignment of the VELO halves, modules, and sensors is determined by allowing their assumed positions to vary a small amount in fits that minimise the mean hit residuals and χ^2 of tracks fitted in the VELO [78, 79, 80]. There are four different VELO alignments, of progressively improving quality, that were developed during the 2010 and 2011 runs. The first alignment is that determined prior to the 2010 run using the initial pre-installation, metrology and beam-absorber collisions; in the second, tracks reconstructed from collisions at LHCb were used to improve the alignment of the two VELO halves; in the third, such tracks were also used to determine the alignment of the VELO sensors and modules; and in the fourth, tracks reconstructed from collisions between the proton beams and residual gas in the LHC beam-pipe ('beam-gas' collisions) were added to those from the p-p collisions to further constrain the sensor and module alignment. As there was an intervention to the downstream trackers between the 2010 and 2011 runs, which may have affected their relative alignment to the VELO, the first three alignments only apply to the 2010 run, and the fourth to the 2011 run.

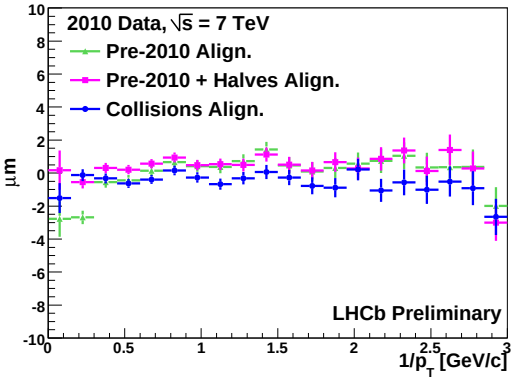
Comparing the distributions of IP resolutions as a function of $1/p_T$ for these different



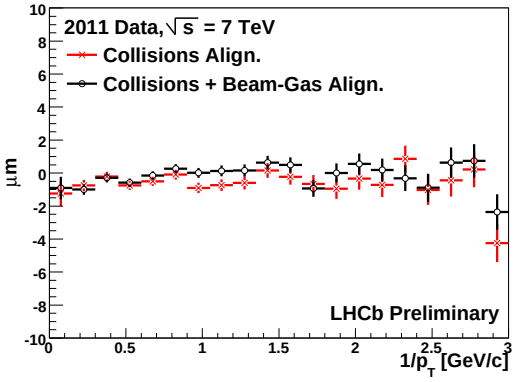
(a) The resolution of IP$_x$ as a function of $1/p_T$ comparing varying qualities of alignment throughout 2010.



(b) The resolution of IP$_x$ as a function of $1/p_T$ comparing two qualities of alignment used in 2011.



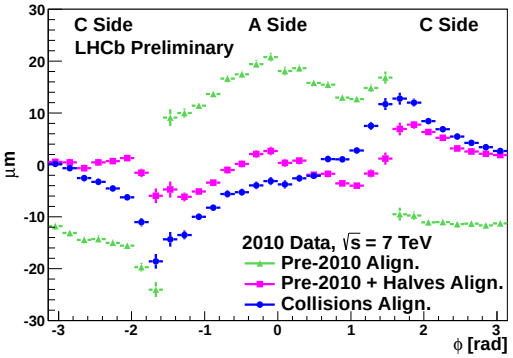
(c) The mean of the Gaussian fits to the IP$_x$ residuals as a function of $1/p_T$ comparing varying qualities of alignment throughout 2010.



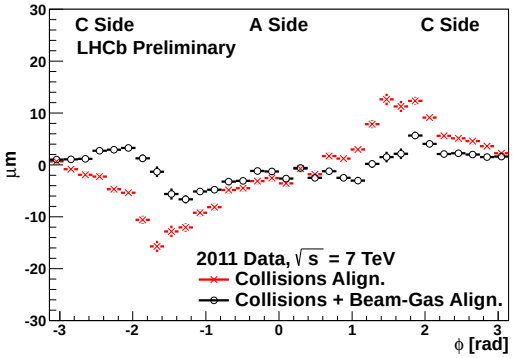
(d) The mean of the Gaussian fits to the IP$_x$ residuals as a function of $1/p_T$ comparing two qualities of alignment used in 2011.

Figure 3.7

alignments provides another method of examining the effect of the VELO alignment on the single hit resolution, and the effect of single hit resolution on IP resolutions. This is shown for 2010 data, using the first three alignments, in figure 3.7a, and for 2011 data, using the 3^{rd} and 4^{th} alignments, in figure 3.7b. As has been shown in section 3.3 the improving alignment, and hence improving single hit resolution, causes the y-intercept of these distributions to reduce, while the gradient remains roughly constant. This is particularly apparent in the difference between the earliest, pre-2010 alignment, and the second alignment, as this marked a significant improvement in the knowledge of the relative positions of the VELO halves. The improvements thereafter are progressively smaller, as the alignment grows closer to the ideal.



(a) The mean of the Gaussian fits to the IP_x residuals as a function of ϕ, comparing varying qualities of alignments throughout 2010.



(b) The mean of the Gaussian fits to the IP_x residuals as a function of ϕ, comparing two qualities of alignment used in 2011.

Figure 3.8

Figures 3.7c and 3.7d show the means of the Gaussian fits to the IP_x distributions for the different alignments as a function of $1/p_T$. No significant deviations from zero are seen, so the IP resolutions appear to be unbiased. However, the p_T of a track has little correlation to any specific region of the VELO, and so is unlikely to reveal any region dependent biases. If, instead, one separates the tracks according to their azimuthal angle ϕ this restricts the tracks analysed to very specific regions of the VELO. In particular, one half of the VELO primarily covers the region $\phi \in (\frac{-\pi}{2}, \frac{\pi}{2})$ (the 'A side'), and the other half the complement of this region (the 'C side'). Figures 3.8a and 3.8b show the mean of IP_x distributions in bins of ϕ, and reveal some significant, ϕ dependent biases.

The pre-2010 alignment shows an offset in the mean IP_x of tracks reconstructed in one half of the VELO to those reconstructed in the other half. This reflects the preliminary nature of the alignment determined with beam-absorber collisions, such that the relative alignment of the VELO halves was not well known. Consequently, IP measurements are also biased, as shown in figure 3.8a, with the sign of the bias depending on the half of the VELO in which the tracks are reconstructed. After re-evaluating the alignment of the VELO halves, in the second alignment, this bias is much reduced, making the mean IP resolutions much more consistent with zero.

The third iteration of the alignment, in which tracks from collisions were used to improve the sensor and module alignment, achieved a general improvement in the single hit resolution. This is seen by the reduction in the y-intercept of the resolution as a function of $1/p_T$ in figure 3.7a. However, when examining the mean of the residuals as a function of ϕ a clear, almost sinusoidal bias is introduced. In general, misalignments to the VELO sen-

sors and modules are random, and so only degrade the overall detector resolution. However, certain parameters in the sensor and module alignment do not affect the single hit residuals when they are varied coherently for a group of (or all) sensors or modules. The alignment process thus has little sensitivity to these parameters, which are known as 'weak modes'. One such weak mode is the rotational alignment of the VELO modules about the beam-line. This must be fixed for the VELO as a whole and the rotational alignment of the modules determined with respect to this. However, tracks reconstructed from p-p collisions originate at the interaction point within the VELO, and so each track only intersects relatively few stations in specific regions of the VELO. This means that different regions of the VELO that are separated in z have little correlation in the alignment algorithms, as very few tracks intersect both regions. This can result in consistent biases to the rotational alignment of the modules in such separate regions. A misalignment to the rotational module positions introduces a bias to the ϕ of tracks reconstructed in the VELO. As $x = r\cos\phi$, a small bias $\delta\phi$ to the measured ϕ results in a bias $\delta x = -r\sin\phi\,\delta\phi$ to the reconstructed x position. This can cause a sinusoidal bias to IP_x as a function of ϕ.

The addition of beam-gas events in the 4^{th} alignment achieved greater constraint on the rotational module alignment. Such collisions are distributed evenly along the beam-line and produce particles that are close to parallel with the beam-line. Hence, their reconstructed tracks can traverse the entire length of the VELO. Collisions of 'satellite' bunches, which consist of protons that are offset by one RF bucket from the main proton bunches and collide at $\pm 700\,\mathrm{mm}$ from the interaction point, were also used. Thus, the rotational alignment of the VELO modules can be better constrained using these data samples. As shown in figure 3.8b, the sinusoidal bias to IP_x is almost completely removed, leaving no significant bias.

Thus, it is clear that examining IP resolutions, and their dependence on the geometric parameters of the tracks used to measure them, provides a sensitive method to evaluate the quality of the VELO sensor alignment. In particular, IP resolutions are sensitive to weak modes in the alignment that cannot be evaluated by examining hit residuals alone.

3.6 Detailed Comparison of Observed Resolutions with Expectations

The preceding sections have shown that IP resolutions depend on many different parameters, most notably: particle momentum, material budget, detector resolution, extrapolation distance to the interaction point, and the detector alignment. The material budget and extrapolation distance to the interaction point vary considerably across different regions of the

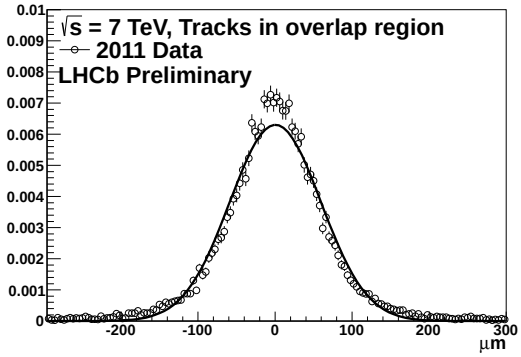
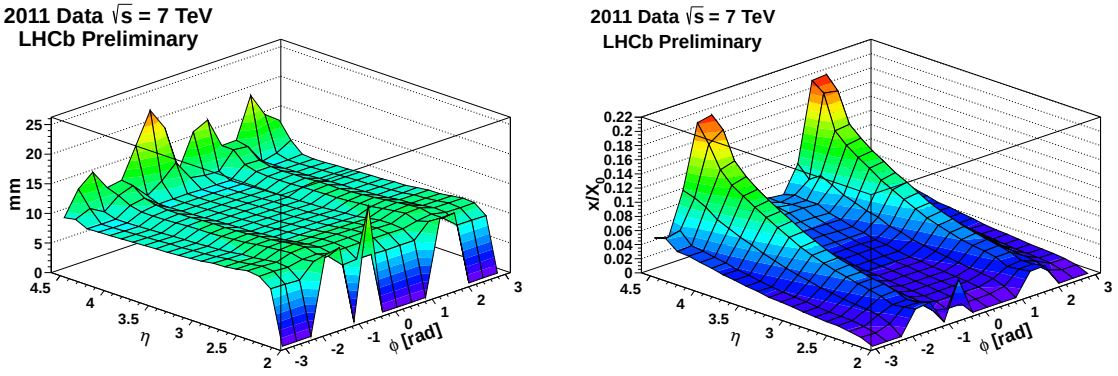


Figure 3.9: The distribution of IP_x residuals for tracks with $1.35 < 1/p_T[\,\mathrm{GeV}/c] < 1.5$ with a single Gaussian fit, as in figure 3.3, using only tracks in the VELO overlap region, defined by $\phi \in \left[\frac{-5\pi}{8}, \frac{-3\pi}{8}\right] \cup \left[\frac{3\pi}{8}, \frac{5\pi}{8}\right]$. Made using 2011 data.

VELO, and simply examining IP resolutions in bins of $1/p_T$ or ϕ is insufficient to separate out these variations. As a consequence the IP distribution in a single bin of $1/p_T$, for example, consists of the integrated contributions from many regions of different material budget, extrapolation distance, *etc.*, and is thus not well described by a single Gaussian, as shown by gigure 3.3. Clearly the single Gaussian only describes the shape very roughly, and fails to fit to either the peak or the tails of the distribution. Figure 3.9 shows the same plot, made using only tracks in the VELO overlap region, as defined in section 3.3. The single Gaussian fit is perhaps slightly improved by the reduced variation in the material budget, but is still quite poor.

Binning finely in η and ϕ splits the VELO into small regions in which the material budget and distance to the interaction point vary very little. To further restrict the variation of the geometric parameters of tracks in any one bin the PVs are required to be within $\pm 50\ \mathrm{mm}$ of the origin.

Figure 3.10 shows the mean values of the parameters of equation 3.12 that contribute to the effects of multiple scattering on IP resolution in bins of η and ϕ. The mean radius of the first hit on a track is fairly constant across the η-ϕ plane. It is slightly larger in the overlap region of the VELO due to the small 'cut aways' at the corners of the VELO sensors, visible in figure 2.7, which cause their sensitive region to start slightly farther from the beam-line. It also decreases slightly at high η as the tracks are closer to parallel with the inside edge of the sensors. The mean material budget intersected by tracks before their first hit is again extracted from the simulated description of the detector, and varies much more. As expected it is considerably larger in the overlap region, due to the overlaps in the RF-foil. It also increases with η, as the tracks intersect the RF-foil at smaller angles and thus travel much farther within it.



(a) **(b)**

Figure 3.10: The mean values of the parameters of equation 3.12 that contribute to the effect of multiple scattering on IP resolutions, in bins of η and ϕ. (a) shows the average radius of the first hit on a track, and (b) the average material budget, in radiation lengths, of the simulated detector intersected by the track before its first hit. The material budget is obtained by extrapolating tracks from real data through the simulated description of the detector, and so may not be the same as the distribution of material in reality. Made using 2011 data.

Figure 3.11 shows the mean values of the parameters of equation 3.12 that contribute to the effects of detector resolution on IP resolution. The resolution of the first hit on a track is estimated from its radius, from which one can obtain the strip pitch and thus estimate the hit resolution. The hit resolution closely follows the distribution of the radius of the first hit, in figure 3.10a, as the strip pitch varies roughly linearly with radial distance from the beam-line. The number of stations, n, used to calculate the extrapolation factor f_{1n} depends on the p of the tracks and the VELO geometry. Tracks with higher p tend to be closer to parallel with the beam-line, and so have larger η, thus n increases with η. However, at very large η the first hit is generally in one of the VELO stations far downstream from the interaction point, and so the number of stations in which a track has hits is limited by the length of the VELO. This is particularly apparent in the overlap region at high η. Here the increased distance to the sensitive region of the VELO sensors means tracks only have hits in the most downstream VELO stations, causing n to decrease. This is reflected in the values of the extrapolation factor f_{1n}. It is slightly larger in the overlap region due to the increased extrapolation distance to the first hit on a track, but is otherwise quite flat for $\eta < 3.8$. This is because the increase of n with η is compensated for by the increase in extrapolation distance to the first hit. At very large η n becomes sufficiently large to cause a significant reduction in f_{1n}, except in the overlap region where the decrease in n and increase in extrapolation distance causes f_{1n} to become large at large η.

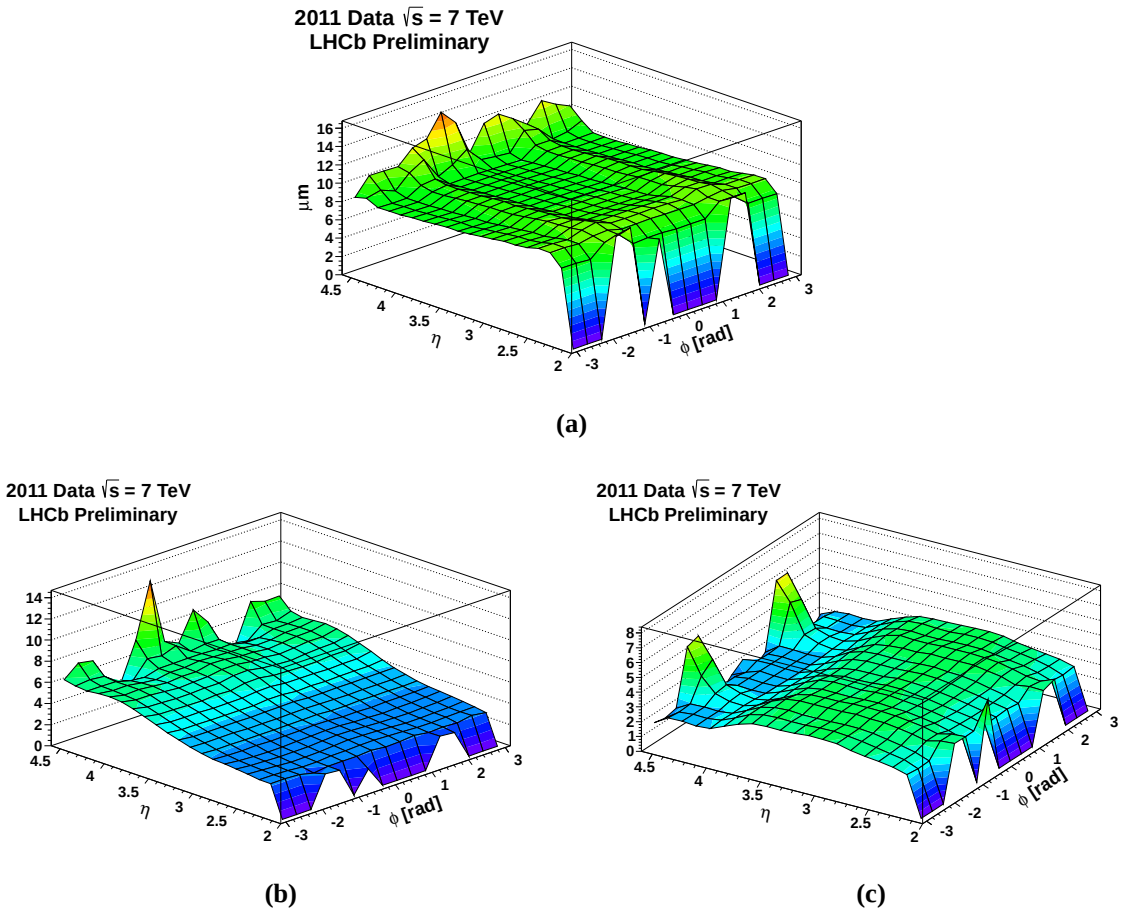


Figure 3.11: The parameters of equation 3.12 that contribute to the effects of detector resolution on IP resolutions, in bins of η and ϕ. (a) shows the average resolution on the position of the first hit on a track, (b) the average number of stations used to calculate the extrapolation factor, f_{1n}, and (c) the average value of f_{1n}. Made using 2011 data.

	a [μm]	b	ϵ [μm]
σ_{PVx}	106 ± 9	0.66 ± 0.07	0.5 ± 1.7
σ_{PVy}	116 ± 12	0.76 ± 0.07	2.6 ± 1.3

Table 3.1: The fitted values of the parameters of equation 3.15, for the x and y PV co-ordinates, taken from the studies in [75].

As mentioned in section 3.2 the measured IP also contains a contribution from the PV resolution, and so the measured IP resolution is in fact

$$\sigma_{IP,meas} = \sqrt{\sigma_{IP}^2 + \sigma_{PV}^2}. \tag{3.14}$$

To properly account for this one must use an independent measurement of the PV resolution and fit the IP distributions with a probability density function (PDF) that includes the PV resolution term. Studies of PV resolution, presented in [75], parametrise it as

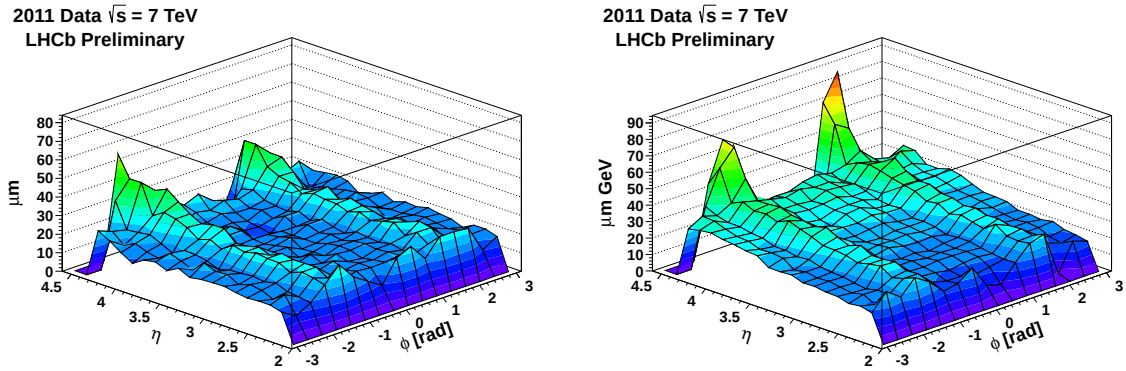
$$\sigma_{PV}(N) = \frac{a}{N^b} + \epsilon, \tag{3.15}$$

where N is the number of tracks used to fit the PV, and a, b and ϵ are fit parameters. The values of a, b and ϵ obtained from the studies in [75] are shown in table 3.1. Equation 3.12 motivates an IP resolution PDF with a constant term and a term dependent on $1/p_T$ summed in quadrature. As N varies for each PV the PDF used for the IP fit is a Gaussian with σ given by

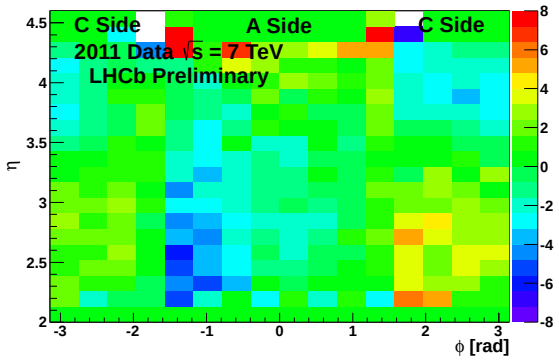
$$\sigma_{IP,meas}(N, p_T)^2 = \sigma_{PV}(N)^2 + A^2 + \left(\frac{B}{p_T}\right)^2, \tag{3.16}$$

where A and B are the fit parameters. An unbinned maximum likelihood fit is then performed to the distributions of IP$_x$ and IP$_y$ to obtain A and B in each bin of η and ϕ.

Figure 3.12 shows the fitted values of A, B and the mean of the Gaussian obtained in each bin of η and ϕ using 2011 data. The values of A follow the distribution of f_{1n} to some extent, though not particularly well at high η. On the other hand the values of B appear to follow the distribution of the material budget quite well. The mean still shows some evidence of the sinusoidal dependence on ϕ caused by the rotational misalignment about the z-axis. The variation of this bias with η shows how it affects different regions of the VELO separately. Nonetheless, it is never more than a few μm from zero, and is thus negligible. Figure 3.13 shows the total fitted PDF, integrated over p_T, η and ϕ, overlaid on the IP$_x$ distribution for 2011 data. There is still some improvement to be made as the very centre of the peak of the distribution is not fitted precisely, but this is sufficient to gain a clear understanding of the behaviour of IP resolutions.



(a) **(b)**



(c)

Figure 3.12: The fitted values of (a) A and (b) B as defined in equation 3.16 and (c) the mean, in μm, of the Gaussian in bins of η and ϕ. From fits to the IP$_x$ resolution using 2011 data.

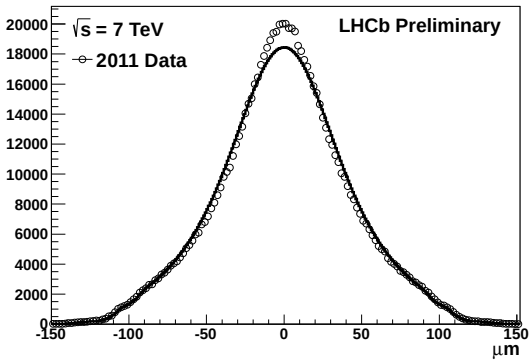
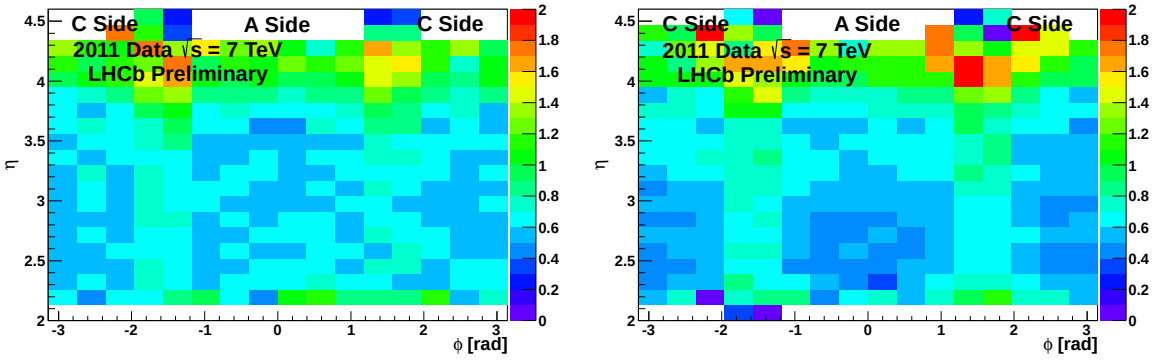


Figure 3.13: The fitted PDF integrated over p_T, η and ϕ, overlaid on the IP$_x$ distribution for 2011 data.



(a) **(b)**

Figure 3.14: The ratio of the measured value of A, as defined in equation 3.16, to that predicted by equation 3.12, for (a) 2011 data and (b) simulated data.

Using the values of the parameters of equation 3.12 shown in figures 3.10 and 3.11 one can compare the values of A and B found with those predicted by equation 3.12. This gives

$$A = f_{1n}\sigma_1, \text{ and} \tag{3.17a}$$

$$B = 0.0136\sqrt{x/X_0}[1 + 0.038\ln(x/X_0)]r_1. \tag{3.17b}$$

Figure 3.14 shows the ratio of the measured values of A to those predicted by equation 3.12 for both 2011 data and simulated data. There is clearly not a very good agreement, though in the region $\eta < 3.8$ the ratio is at least roughly flat outside the overlap region, despite being less than 1. This is likely due to the fact that the parametrisation $n = 2 + \lfloor p/4 \rfloor$ used to calculate f_{1n} assumes that tracks only intersect the VELO sensors after their first hit. This is true outside the overlap region at low η, causing the ratio to be roughly flat. The offset from 1 is likely due to the very rough estimation made of the resolution of the first hit. Elsewhere tracks continue to intersect the RF-foil between VELO stations after their first hit, and so the displacement due to multiple scattering in between VELO stations is significantly larger. This would cause n to be smaller and f_{1n} larger in these regions, and could account for the poor agreement between the observed and predicted values. A parametrisation of n that takes into account variations in the material budget and a more careful extraction of the first hit resolution might make the predicted values more accurate. Nonetheless, the parametrisation of the effects of detector resolution on IP resolution appears to be approximately correct.

Figure 3.15 shows the same for the p_T dependent parameter B. It is particularly striking that on simulated data the ratio is very close to 1 across the η-ϕ plane, showing that the predictions of equation 3.12 are very accurate in this respect. On 2011 data there is a good agreement between the observed and predicted values in the overlap region, but elsewhere

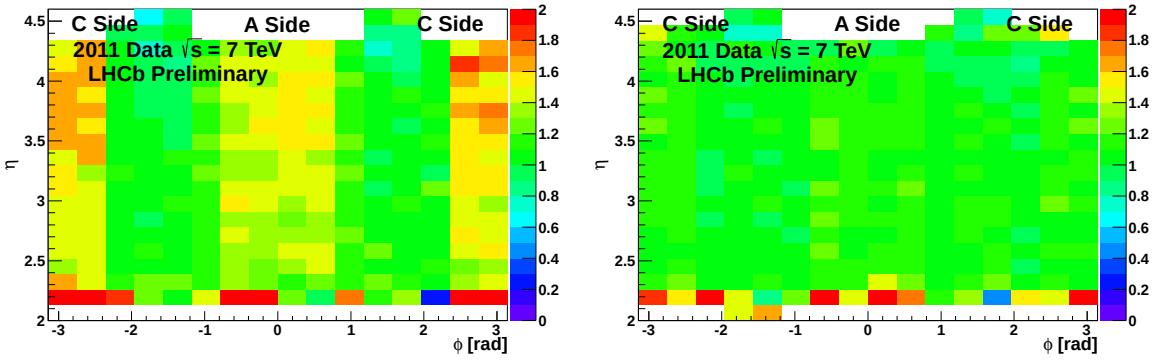
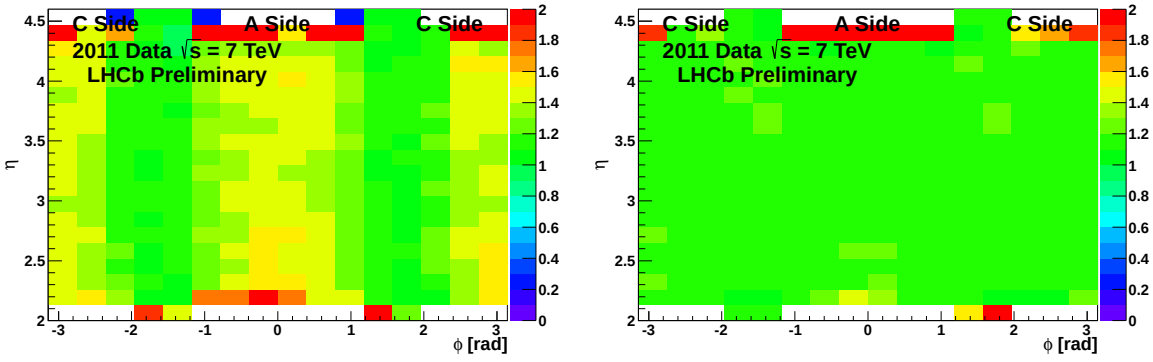


Figure 3.15: The ratio of the measured value of B, as defined in equation 3.16, to that predicted by equation 3.12, for (a) 2011 data and (b) simulated data.

there is a large discrepancy. Given the excellent agreement between prediction and observation on simulated data it appears that the description of multiple scattering in the simulation is accurate, and one is led to conclude that the values of x/X_0 input to equation 3.12 are incorrect for 2011 data in this region.

This can also be seen by examining the 'pull' distributions for the IP measurements. The pull is defined as $(IP_{meas} - IP_{true})/\sigma_{IP}$, where σ_{IP} is the estimated error on the IP calculated by the reconstruction algorithms. As IP_{true} is zero for particles produced exactly at the interaction point the pull is simply IP_{meas}/σ_{IP}. If the errors are estimated correctly the pull distribution should be consistent with a Gaussian with mean zero and $\sigma = 1$. Figure 3.16 shows the σ of Gaussian fits to the pull distribution of IP_x in bins of η and ϕ, for 2011 data and simulated data. For simulated data the σ is consistent with one across the η-ϕ plane, showing that the errors are estimated correctly. For 2011 data the errors are accurate in the overlap region, but are significantly underestimated elsewhere, again showing that the parameters input to the error calculations are incorrect in these regions.

Using the measured values of B and the mean radius of the first hit on a track one can then use equation 3.17b to extract x/X_0. Figure 3.17 shows the extracted values of x/X_0 and the ratio of these to the values extracted from the simulated description of the detector, for 2011 data. The variation in material budget estimated from the values of B is significantly less than that of the simulated detector, as the material budget outside the overlap region is estimated to be $\sim$2 times larger. This is strong evidence that there is material missing from the simulated detector. However, as discussed in section 3.4 and in [74], the simulated description of the VELO is in fact known to be reasonably accurate. Regardless, it is clear that the parametrisation of the effects of multiple scattering on IP resolution is very accurate,



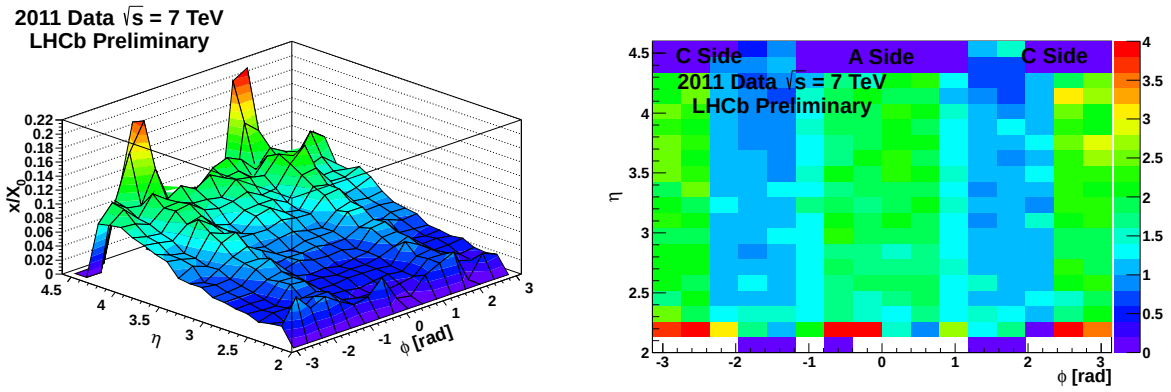
(a) **(b)**

Figure 3.16: The σ of Gaussian fits to the momentum integrated pull distributions of IP_x in bins of η and ϕ. For (a) 2011 data and (b) simulated data.

and that measuring IP resolutions provides a powerful method of examining the distribution of material in the VELO.

3.7 Summary and Conclusions

This section discussed the factors that determine impact parameter (IP) resolution in the VELO. A method of measuring IP resolution without the use of simulation was presented, and its results shown. A mathematical parametrisation of the IP resolution, factoring in contributions from detector geometry and resolution and the effects of multiple scattering was derived in section 3.1. Measurements of IP resolutions on 2011 data and simulated data were presented and compared. In section 3.3 the general behaviour of IP resolutions on hit resolution, detector alignment, extrapolation distance, and material budget was examined and found to be consistent with expectations. A significant discrepancy between IP resolutions on 2011 data and simulated data was also found, and discussed in section 3.4. A more detailed comparison of the predictions of the parametrisation to the measurements made on 2011 and simulated data was also made in section 3.6. The parametrisation of the effects of detector resolution was found to be of limited accuracy, and suggestions were made as to how to improve it. The parametrisation of the effects of multiple scattering were found to be very accurate on simulated data, but less so on 2011 data. This would most logically suggest that there is material missing from the simulated description of the detector. However, these results appear to disagree with complementary analyses made of the VELO material, and so the source of the discrepancy is still under investigation. The most likely source of inaccuracy in the simulated description of the VELO is the RF-foil. Its shape is known to



(a) **(b)**

Figure 3.17: (a) The values of x/X_0 extracted from the p_T dependent term of fits to the IP$_x$ distributions from 2011 data and (b) the ratio of these with the values extracted from the simulated description of the detector.

the simplified in the simulation, and studies are under way into the effects of using a more accurate description of the RF-foil.

Chapter 4

Method of Extracting the Lifetime of a Particle

Both of the parameters y_{CP} and A_Γ, described in section 1.3, involve measurements of the mean proper decay time ('lifetime' for short) of the D^0 meson. The constant decay probability as a function of time of a long lived particle results in its proper decay time following an exponential distribution. Extracting the lifetime from a sample of pure signal candidates reconstructed using an ideal detector would thus be trivial. However, in reality one must deal with backgrounds that can mimic the decay channel of study when reconstructed under the signal hypothesis, particularly in the complex, hadronic environment at LHCb. Candidate selection criteria that are used to exclude backgrounds, both in the trigger during data-taking and offline, can bias the observed lifetime of the signal candidates. Any measurement made by the detector also has a finite resolution. Accounting for these effects makes the extraction of the lifetime rather more challenging. This section describes the various methods used to do so in order to measure y_{CP} and A_Γ.

Section 4.1 describes the essentials of fitting distributions of variables from data with probability density functions in order to find the optimal values of the parameters that determine their shape. The various sources of background and how they can be distinguished from signal are discussed in section 4.2. The methodology required to extract the lifetime of the signal candidates is described in section 4.3, with particular attention paid to the method of correcting for a lifetime biasing candidate selection. The full probability density function required to extract the lifetime is presented in section 4.4. Finally, possible extensions and improvements to the methods presented here are discussed in section 4.5.

4.1 Obtaining the Best Fit to Data

4.1.1 Parameter Optimisation Using the Maximum Likelihood

Any property of the signal candidates in a dataset can be described statistically as a random variable. Such a variable, x, follows a probability density function (PDF), $f(x)$, which gives the probability that its value lies within a certain interval

$$P(x \in [x_1, x_2]) = \int_{x_1}^{x_2} f(x)dx. \tag{4.1}$$

A necessary requirement for a PDF is thus that it must be normalised, $\int_{-\infty}^{+\infty} f(x)dx = 1$, *i.e.* that x must take a finite value for any one candidate in the dataset. This has the corollary that $f(x) \overset{x\to\pm\infty}{\to} 0$. Further, if x is a physical observable, $f(x)$ and its first derivative must be continuous. Otherwise, a PDF can take almost any form. A 'parametric' PDF is one that is described by a specific mathematical formula, the exact shape of which is determined by a set of free parameters. The normalised Gaussian function is an example of a parametric PDF, for which the parameters are the mean and σ.

As a long lived particle has a constant probability of decaying as a function of time its proper-decay-time distribution can be described by an exponential function. The corresponding PDF is thus

$$f(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}} \Theta(t), \tag{4.2}$$

where t is the measured proper decay time, $\Theta(t)$ is the Heaviside step function, which is zero for $t < 0$, and τ is the lifetime. This satisfies $\int_{-\infty}^{+\infty} f(t)dt = 1$ and $\langle t \rangle \equiv \int_{-\infty}^{+\infty} t f(t)dt = \tau$. To extract an unknown τ from a dataset one must determine the value of τ that gives the maximum likelihood for the data. The likelihood function is

$$L(t_1, ..., t_n, \tau) = \prod_{i=1}^{N} f(t_i, \tau), \tag{4.3}$$

where t_i are the proper decay times of each of the N candidates in the dataset. Taking the natural logarithm of the likelihood the product is transformed into a sum while the maximum is still attained at the same value of τ. Substituting in the explicit form of $f(t)$ one obtains

$$\begin{aligned} \ln(L(t_1, ..., t_n, \tau)) &= \Sigma_{i=1}^{N} \ln(f(t_i, \tau)) \\ &= \Sigma_i \ln\left(\frac{1}{\tau} e^{-\frac{t_i}{\tau}} \Theta(t_i)\right) \\ &= -N \ln(\tau) - \Sigma_i \frac{t_i}{\tau}. \end{aligned} \tag{4.4}$$

Differentiating with respect to τ one finds that the maximum is achieved when

$$\begin{aligned} \frac{d}{d\tau}(\ln(L)) &= -\frac{N}{\tau} + \frac{1}{\tau^2}\Sigma_i t_i \\ &= 0, \end{aligned} \tag{4.5}$$

which gives

$$\tau = \frac{1}{N}\Sigma_i t_i. \tag{4.6}$$

So the optimal value of τ is given by the mean of the values of t_i in the dataset, as one would expect. This also means that fitting for the lifetime of a sample of tagged D^0 or $\bar{\mathrm{D}}^0$ using a single exponential PDF will yield the 'effective lifetimes' as described in section 1.3. In section 4.3 it will be shown how this simple PDF needs to be modified in order to be used on a realistic dataset.

The likelihood function can easily be generalised and maximised to find the optimal values of arbitrary parameters of arbitrary PDFs, provided a single solution exists. The values of the parameters thus obtained are those for which the given PDFs best fit the data. It is this method that is used for parameter optimisation in the following sections. The optimisation process itself is performed by the software package MINUIT [81]. The optimisation algorithm varies the values of the free parameters in the fit and examines the value and the rate of change of $\ln(L)$ in order to find the parameter values that yield the maximum $\ln(L)$. The amount by which the parameters are varied is iteratively refined to give a high precision on their optimal values. The statistical uncertainties on these values are determined by finding the variation in the value of each individual parameter that results in a change in the log likelihood of $\Delta \ln(L) = \frac{1}{2}$.

4.1.2 Multi-Dimensional PDFs for Multiple Classes of Candidate

As will be shown in the next section realistic data consists of candidates of several different background classes in addition to true signal candidates. This must be accounted for in the fit PDF. Further, distinguishing these backgrounds from the signal requires examining the distributions of more than one variable. The PDF used to perform such a fit must describe the distributions of all variables of interest for all classes of candidate.

The set of discriminating variables, $\underline{\mathbf{X}}$, is chosen such that each class of candidate follows a different distribution for these variables. This allows a statistical separation of the different classes to be achieved. In this case the PDF required to describe the data becomes the linear sum of the PDFs for each class. The coefficients in this sum are the relative fractions of each class in the dataset. Thus the full PDF is given by

$$f(\underline{\mathbf{X}}) = \Sigma_{class} P(class) f_{class}(\underline{\mathbf{X}}), \tag{4.7}$$

where $P(class)$ is the relative fraction of that class of candidate, and $f_{class}(\underline{\mathbf{X}})$ is the PDF of the variables $\underline{\mathbf{X}}$ for that class. In general the relative fractions of each class will be unknown, and so will be allowed to float in the fit, in addition to the parameters of the PDFs. In this

case one must use the constraint that $\Sigma_{class} P(class) = 1$ to remove one of these fractions from the fit, and allow a unique solution to be found by the optimisation process. Once the optimisation has been performed one can use the PDFs to determine the probability that a given candidate is of a given class using

$$\begin{aligned} P(class|\underline{\mathbf{X}}_i) &= \frac{P(class) f_{class}(\underline{\mathbf{X}}_i)}{f(\underline{\mathbf{X}}_i)} \\ &= \frac{P(class) f_{class}(\underline{\mathbf{X}}_i)}{\Sigma_{class} P(class) f_{class}(\underline{\mathbf{X}}_i)}, \end{aligned} \tag{4.8}$$

where $\underline{\mathbf{X}}_i$ are the values of the variables $\underline{\mathbf{X}}$ for candidate i.

When $\underline{\mathbf{X}}$ consists of more than one variable any correlation between its constituent variables must be taken into account in their PDFs. If $\underline{\mathbf{X}} = \{x, y, z\}$ then the PDF can be factorised as

$$\begin{aligned} f_{class}(x, y, z) &= f_{class,xy}(x, y|z) f_{class,z}(z) \\ &= f_{class,x}(x|y, z) f_{class,y}(y|z) f_{class,z}(z), \end{aligned} \tag{4.9}$$

where $f_{class,y}(y|z)$ is the PDF of the variable y for that class given the value of z (also known as the conditional PDF of y), and so on. If the value of y is independent of the value of z then $f_{class,y}(y|z) = f_{class,y}(y)$. If any of the PDFs in this factorisation are identical for all classes then they factor out of the full PDF, equation 4.7. Consequently the contribution of such a PDF to the log likelihood is independent of the other PDFs, and so can be neglected in the determination of their optimal parameter values. However, if the PDFs are different for any two of the classes of candidate they must be included in the fit PDF in order to obtain the correct optimal values of $P(class)$ and the PDF parameters. This is known as the 'Punzi effect' [82]. For example, this must be taken into account when using a proper-decay-time PDF that uses the per-candidate error on the proper decay time to account for the detector resolution: if the distribution of the per-candidate errors is different for signal and background then the PDF of the per-candidate errors must also be factored into the fit PDF.

If one can factorise the set of discriminating variables $\underline{\mathbf{X}}$ into two independent sets, $\underline{\mathbf{Y}}$ and $\underline{\mathbf{Z}}$, then one can write the fit PDF as

$$\begin{aligned} f(\underline{\mathbf{Y}}, \underline{\mathbf{Z}}) &= \Sigma_{class} f_{class,Y}(\underline{\mathbf{Y}}) f_{class,Z}(\underline{\mathbf{Z}}) P(class) \\ &= \left[\Sigma_{class} f_{class,Y}(\underline{\mathbf{Y}}) P(class)\right] \left[\Sigma_{class} \left(f_{class,Z}(\underline{\mathbf{Z}}) \frac{f_{class,Y}(\underline{\mathbf{Y}}) P(class)}{(\Sigma_{class} f_{class,Y}(\underline{\mathbf{Y}}) P(class))} \right)\right] \\ &= \left[\Sigma_{class} f_{class,Y}(\underline{\mathbf{Y}}) P(class)\right] \left[\Sigma_{class} f_{class,Z}(\underline{\mathbf{Z}}) P(class|\underline{\mathbf{Y}})\right], \end{aligned} \tag{4.10}$$

where $P(class|\underline{\mathbf{Y}})$ is the probability of the given candidate belonging to that class given its values of the variables $\underline{\mathbf{Y}}$, as defined in equation 4.8. The two terms in this PDF contribute

two separate terms to the log likelihood. As the first term here is entirely independent of the second its contribution to the likelihood can be maximised separately, provided the variables $\underline{\mathbf{Y}}$ provide full discrimination between all classes. This allows one to perform two separate fits, each using fewer PDFs, rather than a single fit with many PDFs. The first fit finds the optimal parameters to describe the variables $\underline{\mathbf{Y}}$ and the second fit finds the optimal description of the parameters $\underline{\mathbf{Z}}$, using the PDFs of $\underline{\mathbf{Y}}$ to provide statistical discrimination between the different classes of candidate.

This is useful as performing two fits with fewer free parameters is considerably less CPU intensive than performing one fit with many free parameters. This is because the dimensionality of the likelihood space is given by the number of free parameters in the fit. Thus, the number of steps required to determine the maximum likelihood, *i.e.* the number of different combinations of different parameter values that must be tried by MINUIT, grows exponentially with the number of free parameters. Consequently, treating the PDFs of variables that are known to be independent separately significantly reduces the number of steps required to find their optimal parameter values, and can thus save a great deal of CPU time. However, one must be certain that the variables concerned are indeed fully independent.

4.1.3 Obtaining Non-Parametric PDFs from Data Using Kernel Density Estimation

The PDFs used to describe data often have a physically motivated form, with parameters that have physical interpretations. The exponential PDF used to describe the proper-decay-time distribution of a long lived particle, discussed in section 4.1.1, is one example of this. In some cases there is no physically motivated reason to use any particular shape of PDF to describe the distribution of a given variable. In this case one can use parametric PDFs with sufficient flexibility as to allow them to fit the data. For example, a sum of several exponential PDFs with different τ values might be used to describe the proper-decay-time distribution of combinatorial backgrounds. This has the disadvantage that it introduces a large number of free parameters into the fit, and so can affect the fit stability and significantly increase the fit time. An alternative, preferable option is to obtain the PDF from the data itself.

Simply histogramming a variable is insufficient for use as a fit PDF, as binning a continuous distribution necessarily loses information on its shape and is likely to introduce a bias into the fit. A more accurate way to reproduce a continuous distribution from data is to use kernel density estimation [83]. In this technique each candidate in a dataset contributes one 'kernel function' to the total distribution, rather than just one entry in one bin of a histogram. A kernel function, $K(x)$, must be normalised to have an area of one and have a

mean of zero. They serve to spread the contribution of each candidate over a certain range. The kernel function for a given candidate is translated to have mean equal to the value of the variable of interest for that candidate. The width of the kernel functions is adjusted by introducing a 'bandwidth' parameter h so that the kernel function becomes $\frac{1}{h}K((x-x_i)/h)$, for a candidate with measured variable value x_i. The total distribution is then given by the sum over the candidates in the dataset of their kernel functions

$$f(x) = \frac{1}{Nh}\Sigma_{i=1}^{N} K((x-x_i)/h). \tag{4.11}$$

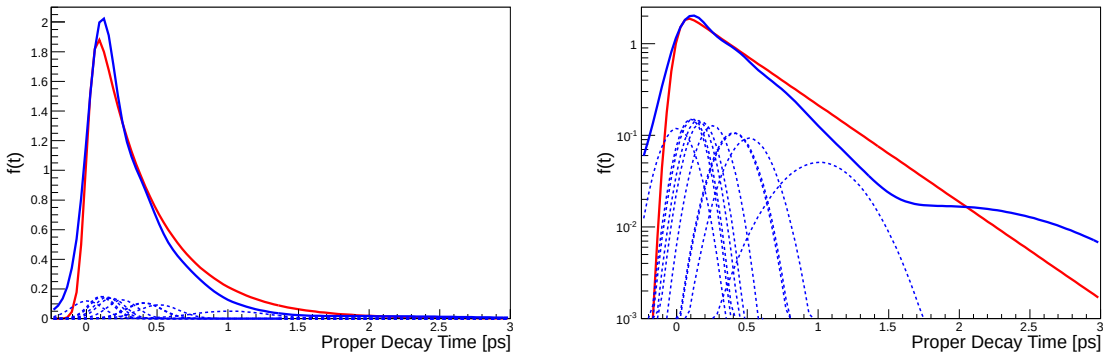
The Gaussian function provides an ideal kernel function as it is continuous and infinitely differentiable. The bandwidth parameter h thus corresponds to the σ of the Gaussian.

The bandwidth determines the amount by which the contribution of each candidate is spread, and is thus comparable to the bin width of a histogram. Using too small a bandwidth results in too little overlap between the contributions of each candidate, and a final distribution that is not smooth, while too large a bandwidth can over-smooth the distribution, and distort its shape. In [83] the optimal bandwidth is derived by minimising the mean integrated squared error of the total distribution, $f(x)$. Further, an adaptive bandwidth, which is different for each candidate, is introduced and found to be

$$h_i = \left(\frac{4}{3N}\right)^{1/5} \sqrt{\frac{\sigma}{f_0(x_i)}}, \tag{4.12}$$

where N is the number of candidates in the dataset, σ is the standard deviation of the variable x, and $f_0(x_i)$ is an a priori estimate of the value of the PDF for the given candidate. Here it is sufficient to use a normalised histogram to estimate $f_0(x)$. This has the properties that larger datasets result in narrower kernels, while regions of lower density (lower $f_0(x)$) result in broader kernels. One can further adapt the bandwidth by using the uncertainty on the measured value for each candidate as the σ parameter, such that regions where the uncertainties are larger result in wider kernels. Alternatively, one can use the 'local standard deviation', *i.e.* the standard deviation calculated using the n candidates with value of x closest to that of the candidate for which the bandwidth is being determined, for some n. This would result in smaller σ in peaking regions and larger σ in flatter regions. Figure 4.1 shows an example of a PDF obtained from toy data using this method with a fixed value of σ.

One issue of this technique is in reproducing distributions of variables that are bounded within a certain range. In this case the kernel function must be adjusted to be defined and normalised only within the range of the variable x. There are two common ways of dealing with this: either to truncate and normalise the kernel function within the given range, or to reflect the kernel function at the boundary. Both these options are problematic as they alter the mean of the kernel function from the desired value, x_i, generally shifting it farther from



(a) **(b)**

Figure 4.1: An example of a proper-decay-time PDF reproduced from data using kernel density estimation, with 50 candidates. (a) and (b) show the same plot using a linear and log scale respectively. The dashed blue lines show the individual Gaussian kernel functions for the first 10 candidates, and the solid blue line the sum of all kernel functions. The solid red line shows the distribution used to generate the proper decay times of the candidates.

the boundary. Such a bias to the mean of the kernels results in slightly lower density very near the boundary, and slightly higher density at a small distance from the boundary, compared to the distribution that the data actually follow. Kernel functions that retain the correct mean in the presence of a boundary do exist, but are unsuitable for use in building a PDF in this manner [83]. Nonetheless, this deviation is small, and kernel density estimation remains a reliable method of obtaining a PDF from data. Any deviation of the PDF obtained using kernel density estimation from the true distribution of the data will result in a systematic uncertainty on the final measurement. The size of this systematic uncertainty can be estimated by varying the σ parameter used to calculate the bandwidth in equation 4.12. This is done for the analyses presented in chapters 5 and 6 in section 6.3, where it is found to be small.

The computational implementation of this technique poses a few challenges. Firstly, evaluating the value of the PDF for a given value of x using equation 4.11 would require evaluating the value of one Gaussian for each candidate in the dataset. This alone would be computationally intensive. Further, as each PDF in a fit is evaluated once for each candidate for each iteration of the fit, the number of calculations required for each fit iteration would depend quadratically on the number of candidates in the dataset. Consequently, this is not a viable option for large datasets. The solution to this is to sample the exact analytical PDF at fine intervals across the range of the variable x, and store these values. One can then interpolate between these values to evaluate the PDF at intermediate values of x. 3^{rd} order spline interpolation performs this task adequately. In regions of low density where the values

of $f(x)$ are close to zero the spline interpolation can give negative values. In this case simple linear interpolation can be used instead, to ensure the values of $f(x)$ remain positive.

In order for the shape of $f(x)$ to be reproduced accurately by the interpolation there must be a large number of sampling points. Generally $\mathcal{O}(1000)$ is found to suffice. The integral of $f(x)$ can then be approximated for a given range of x by summing over the value of $f(x)$ at each sampling point in the range, multiplied by the width of the sampling intervals. Again this can be very computationally intensive when a large number of sampling points is used, particularly if one needs to evaluate the integral repeatedly. In this case one can simplify the calculation by using a cumulative integral function

$$I(x) = \int_{-\infty}^{x} f(x')\mathrm{d}x'. \tag{4.13}$$

Evaluating $I(x)$ at each sampling point in x, storing the values, and interpolating between them allows quick evaluation of the integral of $f(x)$ using

$$\int_{x_1}^{x_2} f(x)\mathrm{d}x = I(x_2) - I(x_1). \tag{4.14}$$

This involves only a small number of calculations, rather than the several hundred that would be required otherwise. As $I(x)$ is a strictly increasing function the interpolation between its sampling points must also be strictly increasing. Consequently, it is easiest to use linear interpolation for $I(x)$.

Kernel density estimation in this form has been applied to other physics analyses [84, 85] and found to work well. Thus, it provides a reliable method of obtaining PDFs from data.

4.2 Discrimination of Signal and Background

The final datasets used to extract physics parameters consist of candidates reconstructed as detailed in section 2.2.8.1, with the candidate selection applied as described in section 5.1. While the candidate selection can be optimised to maximise the signal content of the dataset there are certain backgrounds that cannot be fully excluded. In this case one needs to examine the distributions of the kinematic variables of the signal and backgrounds, and use those that have sufficient separation to achieve a statistical discrimination between the different candidate types.

Should the signal and background candidates follow sufficiently different distributions in proper decay time it would be possible to distinguish them using only that. However, any class of candidate, whether signal or background, will generally follow something akin to an exponential distribution in proper decay time. Thus the distribution of proper decay

time alone is insufficient to separate signal from background, and additional variables must be used. In the majority of cases it is sufficient to use the reconstructed mass of the D^0, $m(\mathrm{D}^0)$, as the discriminating variable. This is particularly convenient as it is independent of the proper decay time for most classes of candidate.

4.2.1 Backgrounds Distinguishable Using Mass Variables

The physical width of the D^0 is very small, so the mass of the D^0 is essentially constant. Thus, the reconstructed $m(\mathrm{D}^0)$ is simply this constant value plus the resolution term, which ideally gives it a Gaussian distribution centred on the mass of the D^0. $m(\mathrm{D}^0)$ is given by the invariant mass of the sum of the momentum 4-vectors of its daughter particles. For a two body decay this is given by

$$\begin{aligned} m(\mathrm{D}^0)^2 &= (E_1 + E_2)^2 - |\vec{p}_1 + \vec{p}_2|^2 \\ &= \left(\sqrt{m_1^2 + |\vec{p}_1|^2} + \sqrt{m_2^2 + |\vec{p}_2|^2}\right)^2 - |\vec{p}_1|^2 - |\vec{p}_2|^2 - 2\vec{p}_1 \cdot \vec{p}_2 \\ &\simeq 2|\vec{p}_1||\vec{p}_2| - 2\vec{p}_1 \cdot \vec{p}_2 \\ &= 2|\vec{p}_1||\vec{p}_2|(1 - \cos\theta), \end{aligned} \tag{4.15}$$

where θ is the opening angle between the two daughter particles, and the approximation $m_n \ll |\vec{p}_n|$ is made. The resolution on the momenta of the daughter particles has $\frac{\sigma(p)}{p} \propto p$, thus the resolution of $m(\mathrm{D}^0)$ is also dependent on the daughter particle p. As the daughter particle p takes a wide range of values, the PDF of the reconstructed $m(\mathrm{D}^0)$ for signal is generally approximated by a sum of several Gaussians with different σ but a common mean. When there are relatively light charged particles, like $\pi^{\pm}$, in the final state these can radiate photons while traversing the detector, and lose momentum. This results in a lower reconstructed mass for the mother D^0. In such cases the mother D^0 mass distribution can be described by a PDF like the 'Crystal Ball' PDF [86]. This is a Gaussian but with a larger lower tail to model candidates for which the daughter particles have lost momentum through photon emission.

Physics backgrounds that must be considered are certain types of decays that can closely resemble signal when reconstructed under the signal hypothesis. The decays $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ and $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ can be such backgrounds for each other. Working under the $\mathrm{K}^+\mathrm{K}^-$ hypothesis, a $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ decay would result in the mass of the D^0 being reconstructed as larger than the true D^0 mass, as the π is assigned the mass of a K. Conversely, working under the $\mathrm{K}^-\pi^+$ hypothesis, a $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ decay would be reconstructed as a low mass D^0, as the K is assigned the mass of a π. Fortunately, due to the excellent momentum resolution provided by the LHCb tracking systems, the resolution on the mass of the D^0 is sufficiently small that

the mass peaks resulting from decays of either of these types reconstructed under either hypothesis do not overlap to any significant degree. One decay that may need to be taken into account in future is that of $\mathrm{D}^0 \to \mathrm{K}^-\pi^+\pi^0$ when working under the $\mathrm{K}^+\mathrm{K}^-$ hypothesis: in this case the π^0 is not reconstructed, while the π^+ is assigned the K mass. This can result in a candidate with reconstructed mass quite close to the true D^0 mass. However, at the precision of the measurements presented in chapter 5, the level of this background is sufficiently small as to be safely neglected.

Combinatorial backgrounds are those candidates that result from random combinations of particles observed in the detector. These do not correspond to the decay products of a single mother particle, but can still happen to have the right kinematic properties to pass the candidate selection. The distribution of $m(\mathrm{D}^0)$ from combinatorial backgrounds tends not to peak, as they do not correspond to a real decay. Thus the PDF of $m(\mathrm{D}^0)$ for combinatorial backgrounds can generally be described by a low order polynomial, normalised within the mass range covered by the data.

In the case of the measurement of A_Γ the flavour of the D^0 at production is tagged by reconstructing the decay $\mathrm{D}^{*+} \to \mathrm{D}^0\pi_s^+$. The charge of the slow pion, π_s, is used to determine whether a D^0 or $\overline{\mathrm{D}}^0$ was produced. Consequently, another type of background that must be considered is 'randomly-tagged' D^0, where the D^0 is reconstructed correctly but is combined with a random π_s to make the D^{*+}. Assuming π^+ and π^- are produced in equal quantities this results in half of such D^0 candidates being assigned the wrong flavour tag. Such backgrounds will have the correct mass for the D^0 and are thus indistinguishable from signal using $m(\mathrm{D}^0)$ alone. However, the mass of the resulting D^{*+} follows the distribution of combinatorial background. Thus, the mass difference $m(\mathrm{D}^{*+}) - m(\mathrm{D}^0) \equiv \Delta m$ can be used to distinguish randomly-tagged D^0, and determine what fraction of the dataset they constitute. Doing so is necessary as treating randomly-tagged D^0 as signal would bias the measured effective lifetime.

Signal, with the D^{*+} correctly reconstructed, will form a peak in Δm about 145.4 MeV, which is the true difference in mass between the D^{*+} and D^0. Similarly to $m(\mathrm{D}^0)$, the PDF of $m(\mathrm{D}^{*+})$, and thus of Δm, can be modelled by a sum of several Gaussians with a common mean for signal. Further, the resolution of $m(\mathrm{D}^{*+})$ and $m(\mathrm{D}^0)$ are correlated, as they both depend on the momenta of the D^0 decay products, so the resulting resolution on Δm is much reduced.

Combinatorial backgrounds will again be relatively evenly distributed in Δm, with no peak formed. The PDF of Δm is restricted to be zero in the region $\Delta m < m(\pi^+)$, as this region is unphysical. Pure combinatorial backgrounds, where the D^0 is also combinatorial, follow the same distribution in Δm as randomly-tagged D^0. Thus, only by examining the

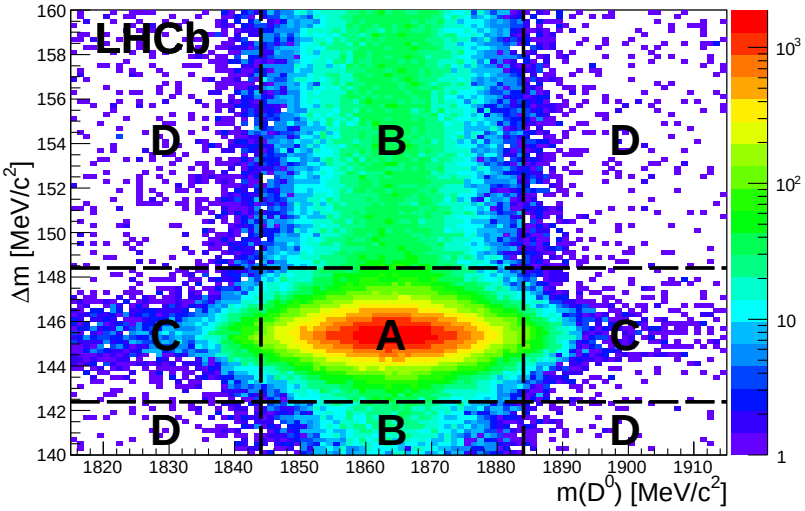


Figure 4.2: The distribution of Δm versus $m(\mathrm{D}^0)$, using 2010 data. Region A covers the signal peaks in Δm and $m(\mathrm{D}^0)$, and contains the majority of the signal candidates. In region B a signal D^0 candidate is combined with a random π_s, so that the D^0 is assigned a random flavour tag; such candidates peak in $m(\mathrm{D}^0)$, but not in Δm, as the D^{*+} is mis-reconstructed. Region C contains the tails of the signal peak in $m(\mathrm{D}^0)$ but also contains a significant fraction of combinatorial backgrounds, while region D is predominantly combinatorial backgrounds. The dataset used here and the selection criteria applied to it are described in section 5.1.

distributions of $m(\mathrm{D}^0)$ and Δm simultaneously can one distinguish randomly-tagged D^0 from pure combinatorial background and signal. Figure 4.2 shows the distribution of Δm versus $m(\mathrm{D}^0)$, and details the types of candidate that dominate each region in the Δm-$m(\mathrm{D}^0)$ plane.

$m(\mathrm{D}^0)$ and Δm are independent by definition, and both are independent of the proper decay time. Thus, if the set of proper decay time variables is $\underline{\mathbf{T}}$ then, following the form of equation 4.10, the fit PDF is given as

$$f(m(\mathrm{D}^0), \Delta m, \underline{\mathbf{T}}) = \left[\Sigma_{class} f_{class,m}(m(\mathrm{D}^0)) f_{class,\Delta m}(\Delta m)\right] \left[\Sigma_{class} f_{class,T}(\underline{\mathbf{T}}) P(class|m(\mathrm{D}^0), \Delta m)\right]. \quad (4.16)$$

4.2.2 Dealing with Secondary D^0

A source of background that contributes in studies of D mesons but not for B mesons is that of D mesons that are produced in decays of B mesons. Approximately 10 % of D^0 observed by LHCb are produced via $\mathrm{B} \to \mathrm{D}^0 X$, where X can be a number of different particles, and the B can be of any species of B meson. As the lifetime of B mesons is $\sim$1.5 ps, and the B is

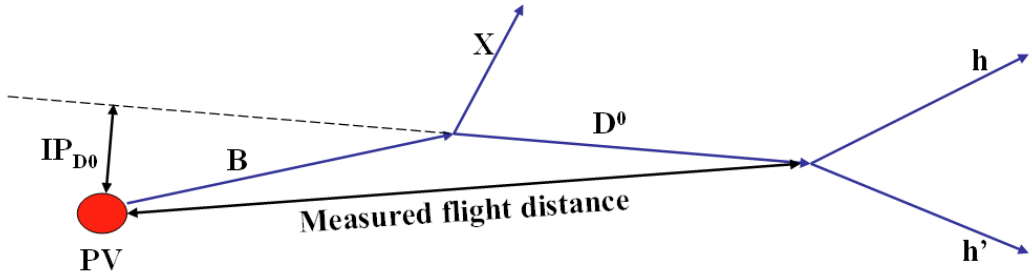


Figure 4.3: A diagram showing how a $\mathrm{B} \to \mathrm{D}^0(hh')X$ decay reconstructed as $\mathrm{D}^0 \to hh'$ results in a positive bias to the D^0 flight distance, and consequently its proper decay time. Also shown is how the impact parameter of such a D^0 is generally larger than that of a particle produced at the PV.

not reconstructed, such decays result in a significantly larger reconstructed proper decay time for the D^0 than is actually the case, as shown in figure 4.3. This would cause a significant positive bias to the measured effective lifetime of the D^0 if these 'secondary' decays are not discriminated from the signal decays, which are produced directly at the primary vertex.

In a flavour tagged sample of D^0 one does have access to the position at which the D^0 is actually produced via the D^{*+} decay vertex, formed by the D^0 and π_s. However, as the mass difference between the D^{*+} and D^0 is only 145.4 MeV and the mass of a π^+ is 139.6 MeV [7] the π_s tends to have very low momentum. As discussed in chapter 3 impact parameter (IP) resolution varies as $\frac{1}{p}$ and so the resolution on the vertex made by the D^0 and π_s is so poor as to make it useless in determining the proper decay time of the D^0. Further, attempting to reconstruct the B, even partially, is non-trivial as its other decay products may be neutral or outwith the LHCb acceptance; this would also create more difficulties when dealing with prompt D^0 decays. Thus it is best to simply reconstruct the D^0 as if it was produced at the PV and calculate its proper decay time accordingly.

This means that secondary D^0 decays cannot be used in measuring the effective lifetimes, and must be statistically discriminated in the fit PDF. Ideally, one would use a variable that is independent of proper decay time to distinguish prompt and secondary D^0, so that their relative fractions are known before the fit to the proper-decay-time distribution is performed. However, the D^0, and D^{*+} for tagged data, are both reconstructed correctly for secondary D^0, and so $m(\mathrm{D}^0)$ and Δm cannot be used to distinguish them. Further, secondary D^0 produced in decays in which the B decays almost immediately are close to indistinguishable from prompt D^0. However, the longer the B lives before decaying the larger the impact parameter of the D^0 is likely to be, as is also shown in figure 4.3. This is because the trajectories of the B and D^0 are not necessarily collinear, and so the larger the distance between the PV and the point at which the D^0 is produced the larger the IP of the D^0 is likely to be. The IP of prompt

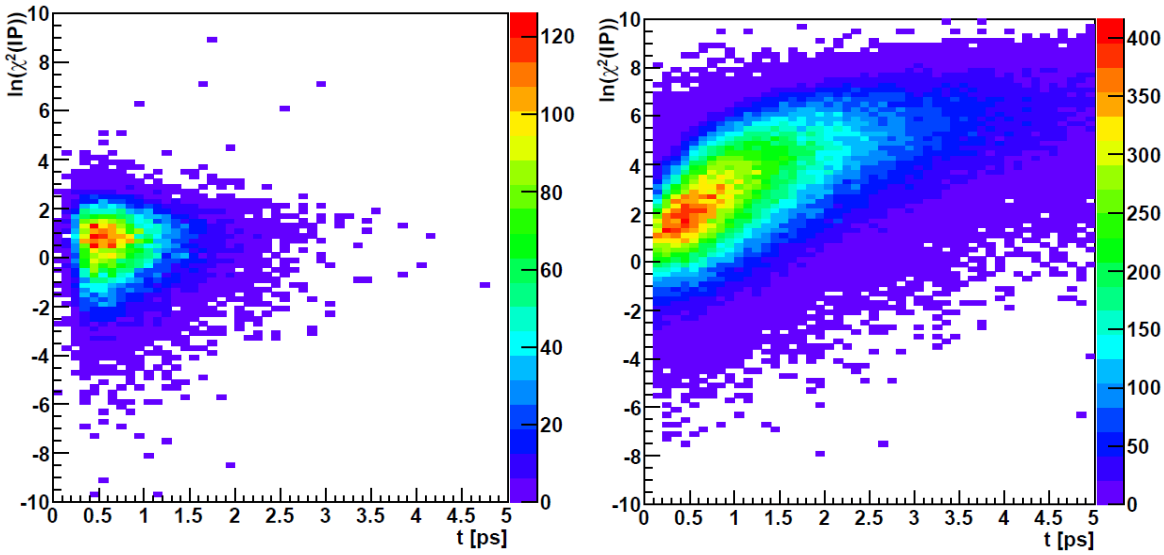


Figure 4.4: The distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ versus proper decay time for (a) prompt D^0 and (b) secondary D^0, obtained from full Monte Carlo simulated data. Reproduced from [87].

D^0 will always be consistent with zero, and is independent of the proper decay time.

Thus one can use the D^0 IP to distinguish prompt and secondary D^0. Given the excellent resolution on IP measurements provided by the VELO it is beneficial to use the $\chi^2(IP_{\mathrm{D}^0})$ as this takes into account the many effects that contribute to the IP resolution. Further, as the $\chi^2(IP_{\mathrm{D}^0})$ can take a very wide range of values for secondary D^0 it is useful to take the natural logarithm, and fit the distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$, which is confined to a much narrower spectrum. Figure 4.4 shows the distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time for prompt and secondary D^0, obtained using full Monte Carlo simulated data. This shows that the distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ is very similar for prompt and secondary D^0 near zero proper decay time, but takes much larger values for secondary D^0 at high proper decay times. As the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF for secondary D^0 is dependent on the proper decay time, the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distribution can only be fitted simultaneously with the proper-decay-time distribution, at the same time as the effective lifetime is determined.

Adding such a PDF to the proper-decay-time part of equation 4.16, and leaving the other variables of the proper-decay-time fit generic as $\underline{\mathbf{T}}$, gives

$$f(m(\mathrm{D}^0), \Delta m, \ln(\chi^2(IP_{\mathrm{D}^0})), \underline{\mathbf{T}}) = \left[\Sigma_{class} f_{class,m}(m(\mathrm{D}^0)) f_{class,\Delta m}(\Delta m)\right] \left[\Sigma_{class} f_{class,T}(\underline{\mathbf{T}}) f_{class,IP}(\ln(\chi^2(IP_{\mathrm{D}^0}))|\underline{\mathbf{T}}) P(class|m(\mathrm{D}^0), \Delta m)\right] . \quad (4.17)$$

Here prompt and secondary D^0 must be treated as a single signal class in the mass component

of the fit, as they are indistinguishable using the mass distributions. The fraction of signal that is prompt (or secondary) is then determined in the proper-decay-time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ fit. For prompt signal one has that $f_{class,IP}(\ln(\chi^2(IP_{\mathrm{D}^0}))|\underline{\mathbf{T}}) = f_{class,IP}(\ln(\chi^2(IP_{\mathrm{D}^0})))$, as the $\chi^2(IP_{\mathrm{D}^0})$ is independent of proper decay time, as shown in figure 4.4a. However, this is likely not the case for combinatorial backgrounds, for which the reconstructed D^0 need not point to the PV. Thus both secondary D^0 and combinatorial backgrounds require a PDF for $\ln(\chi^2(IP_{\mathrm{D}^0}))$ that depends on the proper decay time of the candidate.

The exact form of a PDF for $\ln(\chi^2(IP_{\mathrm{D}^0}))$ is difficult to determine. If the errors on the IP are estimated perfectly then the distribution of $\chi^2(IP_{\mathrm{D}^0})$ for prompt D^0 would follow a χ^2 distribution with two degrees of freedom. However, as was shown in chapter 3, the errors on IP measurements are known not to be estimated correctly. Further, they are unlikely to ever be estimated perfectly, as this would require a perfect parametrisation of the detector resolutions, and perfect simulation of the detector material. So a perfect χ^2 PDF is unlikely to provide an adequate description of the data. Consequently, the approach used in chapter 5 is to use a parametric PDF with a sufficiently large number of free parameters as to allow it to fit the data, for both prompt and secondary D^0. This is described in detail in section 5.3.1.

4.3 Extraction of the Lifetime

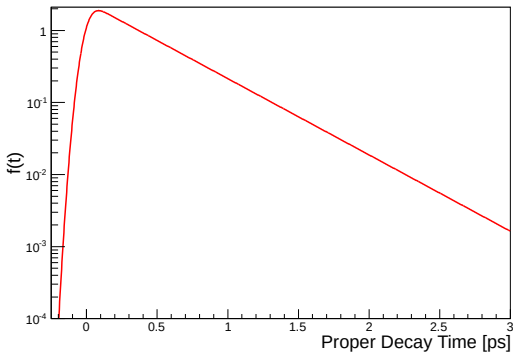
4.3.1 The Proper-Decay-Time PDF

With the backgrounds fully discriminated one can fit the proper-decay-time distribution to obtain the effective lifetime of the signal. Each class of candidates also requires a PDF for the proper-decay-time distribution, to be used in equation 4.17.

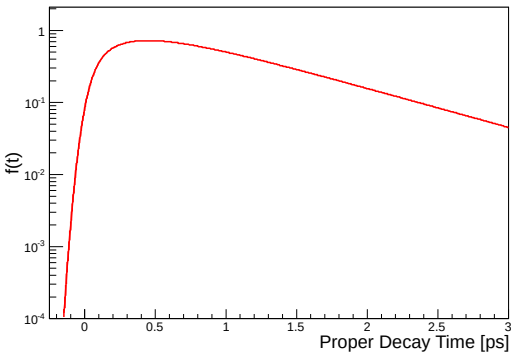
The first consideration that has to be made is that the measurement of proper decay time provided by the detector has an uncertainty on it. The excellent performance of the VELO results in a small proper-decay-time resolution for LHCb, $\sim$50 fs [88], but it must still be taken into account in order to correctly extract particle lifetimes. The measured proper decay time can be described as the true proper decay time of the particle plus a resolution term

$$t_{meas} = t_{true} + \delta t. \tag{4.18}$$

Here the true proper decay time, t_{true}, follows the original distribution, *i.e.* an exponential for signal, while the resolution term, δt, should follow a Gaussian with mean zero and σ equal to the average uncertainty on the measured proper decay time. Thus the measured proper decay time, t_{meas}, can be described for signal by the convolution of an exponential distribution with



(a) An example of the proper-decay-time PDF for prompt D^0, as defined in equation 4.19, with $\tau = 410.1\,\mathrm{fs}$ and $\sigma_t = 50\,\mathrm{fs}$.



(b) An example of the proper-decay-time PDF for secondary D^0, as defined in equation 4.21, with $\tau_1 = 270\,\mathrm{fs}$, $\tau_2 = 800\,\mathrm{fs}$, and $\sigma_t = 50\,\mathrm{fs}$.

Figure 4.5

a Gaussian:

$$\begin{aligned} f(t_{meas}) &= \frac{1}{\tau} e^{-t_{meas}/\tau} \Theta(t_{meas}) \otimes \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{1}{2} t_{meas}^2/{\sigma_t}^2} \\ &= \int_{-\infty}^{\infty} \frac{1}{\tau} e^{-t'/\tau} \Theta(t') \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{1}{2}(t_{meas}-t')^2/{\sigma_t}^2} \mathrm{d}t' \\ &= \frac{1}{\tau} e^{-t_{meas}/\tau} e^{\frac{1}{2}\sigma_t^2/\tau^2} F\left(\frac{t_{meas}}{\sigma_t} - \frac{\sigma_t}{\tau}\right), \end{aligned} \tag{4.19}$$

[84] where the Gaussian frequency function F is defined as

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}y^2} \mathrm{d}y. \tag{4.20}$$

An example of the shape of this PDF is shown in figure 4.5a, where one sees that the resolution effect allows a small fraction of candidates to have negative proper decay times. This PDF would be appropriate for the prompt signal, and the fitted value of τ would yield the effective lifetime as desired. Similarly, the proper decay time of randomly-tagged D^0 can also be modelled using such a PDF, in which case the fitted τ would yield the average effective lifetimes of the D^0 and $\overline{\mathrm{D}}^0$.

The PDF for secondary D^0 is rather more complicated - here one has the contribution from the B flight distance to take into account. This is further complicated by the fact that several different species of B, each with slightly different lifetimes, can produce D^0, and also that the B and D^0 need not necessarily be collinear. Thus deriving an exact, parametric PDF for the secondary D^0 proper-decay-time distribution is very difficult. However, making the simplified assumption that the contribution from the B to the reconstructed D^0 proper decay

time can be modelled by a single exponential, the proper decay time is given by the sum of two exponentially distributed variables. Thus, the PDF is given by the convolution of two exponentials

$$\begin{aligned} f_{sec,t}(t) &= \frac{1}{\tau_1} e^{-t/\tau_1}\Theta(t) \otimes \frac{1}{\tau_2} e^{-t/\tau_2}\Theta(t) \\ &= \int_{-\infty}^{t} \frac{1}{\tau_1} e^{-t'/\tau_1}\Theta(t') \frac{1}{\tau_2} e^{-(t-t')/\tau_2}\Theta(t-t') \\ &= \begin{cases} \frac{1}{\tau_2-\tau_1}\left(e^{-t/\tau_2} - e^{-t/\tau_1}\right) & \text{if } \tau_1 \neq \tau_2, \\ \frac{t}{\tau_1^2} e^{-t/\tau_1} & \text{if } \tau_1 = \tau_2. \end{cases} \end{aligned} \tag{4.21}$$

Accounting for the detector resolution is done analogously to the prompt signal case. In performing a convolution with the Gaussian resolution function the two exponential terms in the $\tau_1 \neq \tau_2$ case yield terms like $\frac{1}{\tau}\times$ (equation 4.19). An example of this PDF is shown in figure 4.5b.

In the ideal case of a single species of B, with the B and D^0 collinear, one τ parameter would correspond to the D^0 lifetime and the other to the B lifetime scaled by $\sim m(\mathrm{D}^0)/m(\mathrm{B})$. However, as the realistic case is more complicated than this simplified model the fitted τ values lose their physical meaning, and simply serve to describe the average shape of the proper-decay-time distribution for secondary D^0. Figure 4.6 shows a fit using this PDF to the proper-decay-time distribution of a sample of secondary D^0 obtained from full Monte Carlo simulation. The PDF describes the data well, showing that the approximations made are good.

Combinatorial backgrounds pose an equally challenging problem in modelling their proper-decay-time distribution, as no physical rationale can be applied to deciding what shape to use. There are various approaches to modelling this distribution. A common method is to use a sum of several exponential PDFs with different τ values. However this has the drawback that, as the size of the dataset increases, more exponential PDFs generally need to be added to the sum in order to obtain a sufficiently accurate description of the distribution. Otherwise a large systematic uncertainty on the measured signal lifetime can result. An alternative approach is to use Monte Carlo simulations of the combinatorial background to obtain the shape of the proper-decay-time distribution, but this requires a great deal of faith in the accuracy of the simulation and can also result in significant systematic uncertainties. In [84] an approach is detailed as to how to obtain a non-parametric PDF using kernel density estimation, while [89] proposes a method in which the information from the mass fit can be used to statistically cancel the contribution of backgrounds to the proper-decay-time fit. However, combinatorial backgrounds are in fact neglected in the fits performed to obtain the measurements presented in chapter 5, for the reasons detailed in section 5.3.1, and so will not be

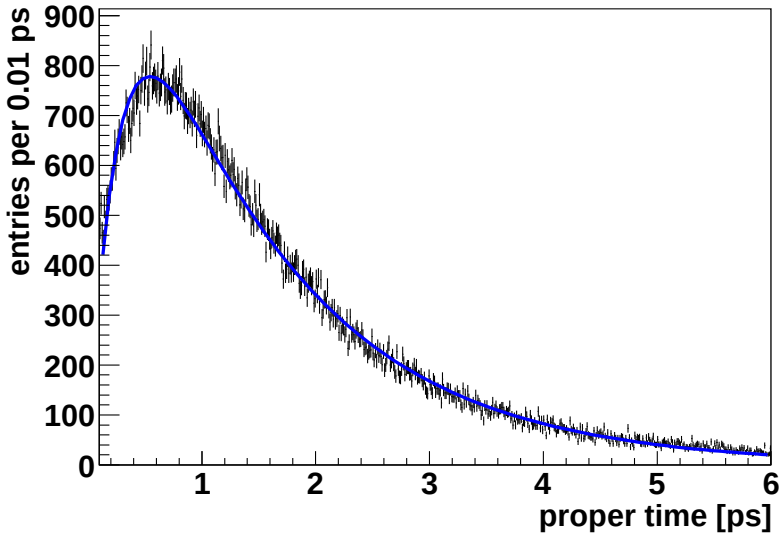


Figure 4.6: The distribution of the proper decay time of secondary D^0, obtained from full Monte Carlo simulated data, fitted with the convolution of two exponential functions and a Gaussian function.

discussed any further here. However, some discussion of how to go about including such backgrounds in the fit PDF can be found in section 4.5.

4.3.2 Accounting For Lifetime Biasing Candidate Selections

The PDFs presented for prompt and secondary D^0 in the previous section make the assumption that the selection criteria applied to the final data have no dependence on the proper decay time of the candidates. In this case the distributions are unmodified from the ideal case beyond the effects of detector resolution. However, as was discussed in chapter 3, the vast majority of particles detected in LHCb are produced directly at the PV. This can result in an overwhelming quantity of combinatorial background when searching for relatively rare decays.

Such background candidates will for the most part have very low proper decay times, as their constituent particles are prompt. Thus it is often beneficial to apply selection criteria that favour longer lived candidates, which are much more likely to correspond to a long lived particle, and so exclude much of the combinatorial background. However, this also rejects any short lived signal, and modifies the observed proper-decay-time distribution from its original shape. Should the specific decay examined have a particularly unique characteristic, such as a pair of muons that can be identified very cleanly in the final state, or a resonant intermediate state with very small resolution on its reconstructed mass, such 'lifetime biasing' selection criteria can be avoided. The only way to do this for the decays $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$

and $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ would be to use the PID information provided by the RICH detectors to apply a tight lower bound on the PID to select the Ks. However, the RICH reconstruction is too slow to be used during data-taking in the trigger, and so this is not currently an option. Neural networks and multi-variate selections, using only variables that are independent of proper decay time, can also achieve similar results, but often result in relatively low selection efficiencies. Consequently, the analyses presented in chapters 5 and 6 use lifetime biasing selection criteria.

This introduces a proper-decay-time dependent selection efficiency, which tends to be small at low proper decay times and much larger at high proper decay times. Figure 4.7 shows the proper-decay-time distribution of untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ candidates from 2010 data compared to the expected exponential distribution. The distribution of the data is clearly strongly biased from the original exponential, with very few candidates at low proper decay times. The tail of the distribution also tends to zero more quickly than the unbiased exponential, showing that long lived candidates are also somewhat disfavoured. The reasons for this are discussed in the following section. Common approaches to dealing with this bias are to either use a parametric description of the acceptance as a function of proper decay time, or to obtain the acceptance function from simulations. One can then simply multiply the unbiased PDF by this acceptance function. However, both these approaches tend to introduce large systematic uncertainties due to inaccuracies in the acceptance functions thus obtained. Consequently they are undesirable options when attempting to measure lifetimes to the accuracy required to make competitive measurements of y_{CP} and A_Γ. As y_{CP} and A_Γ only require the calculation of the ratio of effective lifetimes such acceptance effects can be cancelled in the calculation, as is done for $\mathrm{B}^0_s \to \mathrm{K}^+\mathrm{K}^-$ in [85]. However, this requires that the acceptance as a function of proper decay time is the same for all decay modes, or that any differences between them are corrected. The analyses presented in chapters 5 and 6 use a more sophisticated method by which the acceptance as a function of proper decay time is obtained for each individual candidate in the dataset, and then used in the fit.

4.3.2.1 Method of Calculating Per-Candidate Proper-Decay-Time Acceptance Functions

Lifetime biasing selection criteria involve putting limits ('cuts') on any kinematic variables of a candidate that are correlated to its proper decay time. The most common such cut to apply is a minimum limit on the IP, or $\chi^2(IP)$, of the decay products of the candidate. This is very effective at excluding prompt backgrounds, but obviously favours long lived signal candidates. Other lifetime biasing cuts include minimum limits on the flight distance (or its χ^2) of the mother particle, or even simply its proper decay time. Requiring the 'pointing

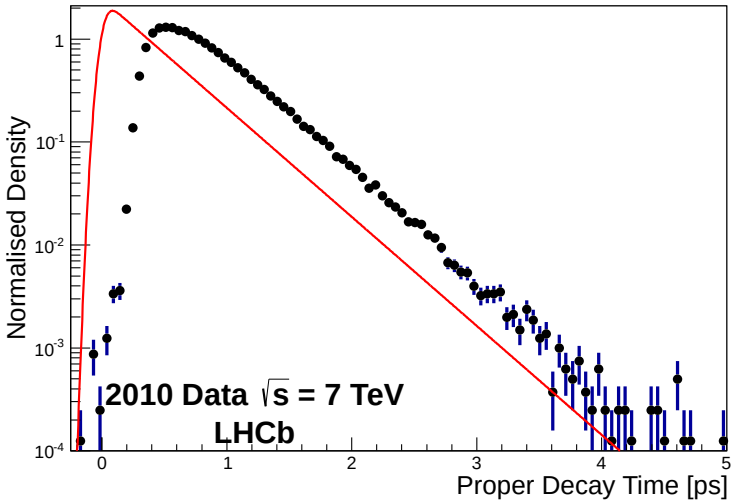


Figure 4.7: The proper-decay-time distribution of untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ candidates using 2010 data (black points) compared to the unbiased exponential distribution expected for signal (red line). The candidate selection clearly disfavours short lived D^0 and biases the observed proper-decay-time distribution. Very long lived candidates are also somewhat disfavoured. The data are selected by requiring $1848 < m(\mathrm{D}^0)[\mathrm{MeV}] < 1880$ and $\chi^2(IP_{\mathrm{D}^0}) < 2.7$, in addition to the selection criteria detailed in section 5.1, to maximise the signal content.

angle', defined as the angle between the momentum vector of the mother particle and the line between its associated PV and its decay vertex, to be near zero also disfavours short lived candidates.

The method to correct for these effects, known as the 'swimming' algorithm, was first used by the NA11 experiment [90], further developed at DELPHI [91] and brought into a form close to that applied here at CDF [92, 93, 94]. It was then implemented at LHCb [84, 85, 95], where the implementation of the high level trigger (HLT) entirely in software makes it particularly applicable, as explained below. The swimming algorithm exploits the fact that the proper decay time of the mother particle is independent of the kinematics of its daughter particles. This allows one to artificially change the proper decay time of the mother particle, by moving its production or decay vertex along the direction of the mother particle's momentum, without needing to modify the daughter particles' kinematics. One can then re-evaluate the decision of the candidate selection for each proper decay time, and thus build the acceptance function for each candidate. An example of how this is done, assuming cuts are placed on the IP of the daughter particles, is shown in figure 4.8.

At each proper decay time the given candidate will either pass or fail the selection, giving acceptance one or zero. Thus the acceptance function can be expressed as a sum of top-hat and step functions. The proper decay times at which the acceptance value changes are referred to as 'turning points', and the acceptance function can be expressed as a vector of these turning points. The proper decay time at which the candidate is actually found naturally falls within a region in which the acceptance is one. Upper 'turn-off' points, at higher proper decay times, can result from applying cuts to the maximum IP of the daughters, or the mother particle's flight distance χ^2, *etc.*, however such cuts are not used at LHCb. Another way in which this can occur is through the use of minimum IP cuts on the daughters in the presence of multiple PVs in the event, as shown in figure 4.9. This is because the IP of any track is taken with respect to all PVs in the event and the minimum value found used as the selection variable. This effect causes a lower acceptance rate on average at high proper decay times.

The most accurate method of changing the proper decay time of a candidate would be to move the mother particle's decay vertex in the direction of the mother's momentum then extrapolate the daughter tracks through the detector to determine where hits in the detector's sensors would be found. The full reconstruction would then need to be run again using these displaced hits to reconstruct the candidate. However, correctly extrapolating the daughter trajectories would be very complicated, and re-running the reconstruction at every step prohibitively CPU intensive. Thus the simpler approach of moving the PVs in the direction of the mother particle's momentum, while keeping the daughter tracks fixed in their measured positions, is taken. This has the same effect of modifying the mother particle's proper decay

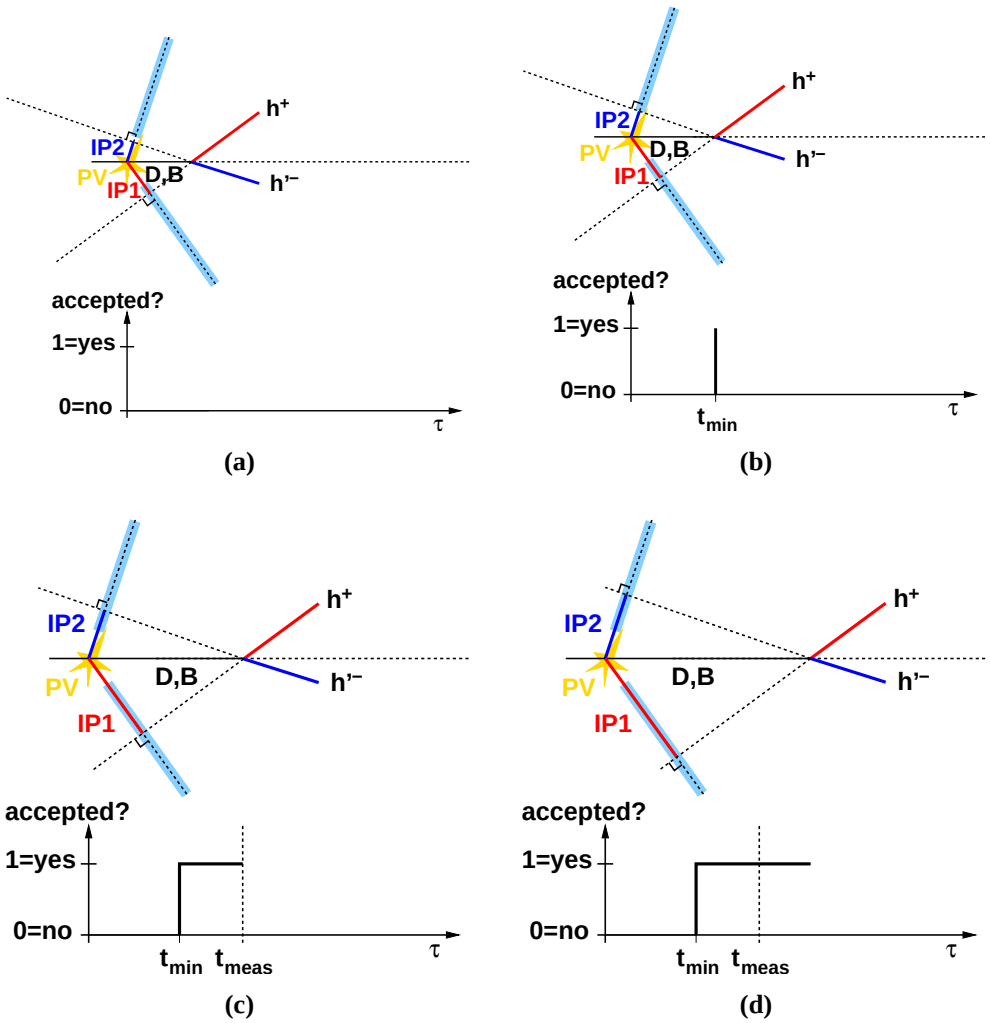


Figure 4.8: An example of how the swimming algorithm determines the acceptance as a function of proper decay time for a given two body decay candidate. Minimum IP cuts are placed on the daughter particles, with the accepted range shown by the blue bands perpendicular to the daughter particles' tracks. In (a) the IP of the positively charged daughter particle is too small and the candidate would fail the selection, giving zero acceptance; at (b) the daughter particle's IP becomes large enough to pass the selection and the acceptance is one; (c) shows the candidate at its measured proper decay time, which is necessarily at a position where the acceptance is one; finally, the algorithm continues to higher proper decay times in (d).

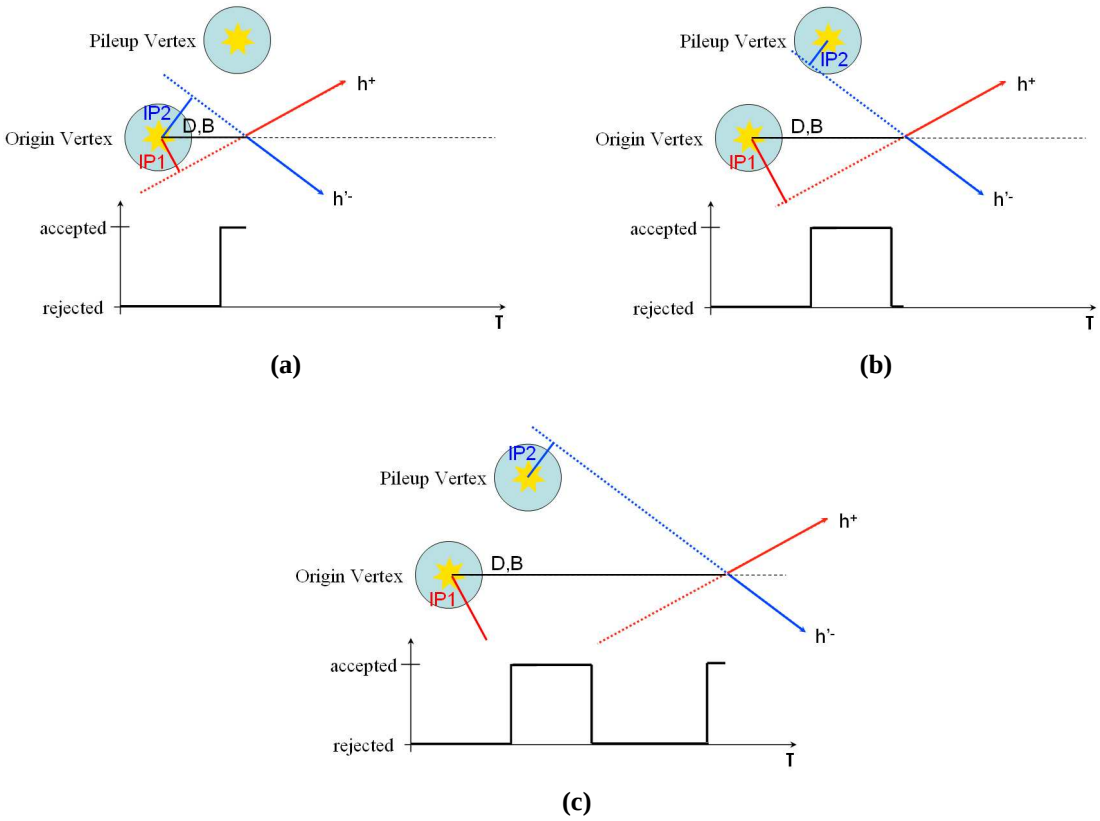


Figure 4.9: An example showing how use of minimum IP cuts on the daughter particles (shown by the blue circles) can cause candidates not to be accepted at high proper decay times when there is more than one PV in an event. In (a) the IPs of both daughter particles are large enough to pass the cuts; in (b) the minimum IP of the negatively charged daughter particle is that taken w.r.t. the pileup PV, and is too small for the candidate to be selected; finally, in (c) both IPs are once again large enough for the candidate to be accepted.

time while being much less prone to error and much more viable in terms of the CPU time required.

The approximation of moving the PVs in the event means that this implementation of the swimming method is insensitive to any variation in the reconstruction efficiency as a function of proper decay time. A lower reconstruction efficiency can occur for mother particles that travel a large distance in the transverse plane before decaying, as the pattern recognition in the reconstruction looks for tracks coming from the beam-line. Such effects are studied in section 6.3.7 and found to be negligible at the precision of the results presented in chapters 5 and 6, but will need to be studied in more detail for future higher precision measurements.

One effect that can be taken into account is that, due to the finite size of the VELO, at very high proper decay times the daughters of the D^0 will no longer have sufficiently many hits in the VELO as to be reconstructable. In the offline reconstruction at least three space points are required to reconstruct a track in the VELO, while in HLT1 this is increased to five. This also naturally causes an upper acceptance effect, though only at very high proper decay times, as a D^0 must live for several tens of times the average lifetime of the D^0 in order for its daughters to intersect so few VELO stations. Nonetheless, this can readily be accounted for by simply checking how many VELO stations the daughter tracks would intersect at each proper decay time had the decay vertex been moved instead of the PVs. The proper decay time beyond which the daughter tracks intersect too few VELO stations becomes the maximum turn-off point. This is then combined with those found for the candidate selection by ensuring that the acceptance is zero beyond this point.

As was described in sections 2.2.7 and 2.2.8 the candidate selection in fact consists of several consecutive stages: the hardware L0 trigger, followed by the software HLT1 and HLT2, and finally the offline stripping selection. The L0 stage triggers only on p_T and E_T, which are independent of proper decay time. Thus the swimming algorithm need not be applied to the L0 trigger. Lifetime biasing cuts are applied in the HLT, but the fact that these are implemented in software allows their specific reconstruction and selection algorithms to be re-run identically as was done at the time of data-taking. This is crucial to the swimming algorithm as it requires re-evaluating the trigger decision at many different proper decay times; any difference between the triggers used when the data are taken and those used for the swimming algorithm would result in large systematic uncertainties on the turning points. Candidates are also required to have triggered the event themselves, and not simply be part of an event that was triggered by some other characteristic. Thus, generally specific trigger (and stripping) lines are selected and applied to the candidates during the swimming algorithm, so that the acceptance functions obtained are with respect to those specific selections. As the HLT and offline reconstruction use different algorithms the swimming algorithm is run on

each separately, and the acceptance functions obtained merged by taking a logical AND.

The specifics of the swimming algorithm are detailed in section 2.4 of [87]. It consists of firstly performing a scan by moving the PVs over a range of $\pm 600\ \mathrm{mm}$ to determine the approximate positions of any turning points, and then refining their positions using iteratively finer, localised scans about the turning points. This could be repeated to obtain an arbitrary precision on the turning point positions, but to limit CPU time a resolution on the turning points of $\sim$4.6 $\mu\mathrm{m}$ is used. This translates into a resolution on the proper decay time of the turning points of $\sim$0.5 fs. There are also various ways in which a bias can enter into the determination of the turning points. These include imperfect reproduction of the trigger decisions, as previously mentioned, as well as differences in the proper-decay-time resolution between the trigger and offline reconstruction. The studies made to evaluate the significance of these effects are also described in [87], where they are found to be negligible.

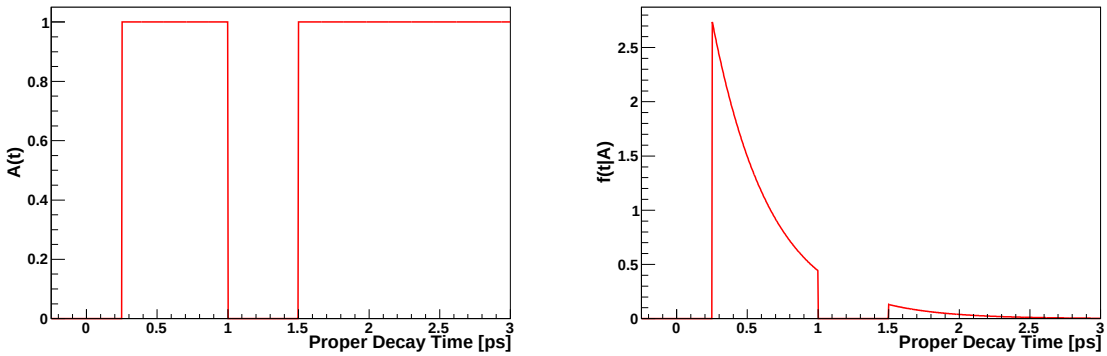
One other complicating effect can occur if not all of the daughters are required to trigger on a specific line. For instance, the HLT1 '1 track' line requires only a single track with high IP to trigger the event. This can be either or both of the D^0 daughters in a two body decay. Further, whether either track triggers the event varies with proper decay time independently for each track, resulting in a much more complicated acceptance function. To avoid this issue only candidates for which both daughters have triggered the event are used in the analyses presented in chapters 5 and 6, which results in excluding 5% of the available candidates.

4.3.2.2 Incorporating Acceptance Function Variables into the Fit PDF

The turning points obtained by the swimming algorithm for each candidate enter into the fit as additional variables in the PDF. Thus the set of proper decay time variables, left generic in equation 4.17, becomes $\underline{\mathbf{T}} = \{t, nTP, TP_1, ..., TP_n\}$, where t is the proper decay time, nTP the number of turning points found, and $TP_{1..n}$ the proper decay times of the turning points. The proper decay time component of the fit PDF thus becomes

$$\begin{aligned} f_{class,T}(\underline{\mathbf{T}}) &= f_{class,T}(t, nTP, TP_1, ..., TP_n) \\ &= f_{class,t}(t|nTP, TP_1, ..., TP_n) f_{class,TPs}(nTP, TP_1, ..., TP_n). \end{aligned} \tag{4.22}$$

As the first turning point is always a turn-on point, and the VELO geometric acceptance results in the last turning point always being a turn-off point, there should always be an even number of turning points. Due to the implementation of the swimming algorithm this is not always the case, as in a few rare instances the scan for turning points does not go far enough to reach the point at which the daughters cannot be reconstructed in the VELO. In this case the final turn-off point is taken to be at $+\infty$. Thus pairing the turning points consecutively yields the intervals in which the candidate would be accepted. The acceptance as a function



(a) An example of the acceptance as a function of proper decay time for a single candidate determined by the swimming algorithm, showing 3 turning points.

(b) The proper-decay-time PDF for prompt D^0 resulting from the acceptance function in (a), as defined in equation 4.24.

Figure 4.10

of proper decay time for a given candidate can thus be written as a sum of top-hat functions:

$$A(t|nTP, TP_1, ..., TP_n) = \Sigma_{n=1}^{nTP/2} \Theta(t - TP_{2n-1})\Theta(TP_{2n} - t). \tag{4.23}$$

The conditional PDF of the proper decay time given the turning points is simply the unbiased proper-decay-time PDF, like those discussed in section 4.3.1, restricted to and normalised in the range in which the candidate would be accepted:

$$f_{class,t}(t|nTP, TP_1, ..., TP_n) = \frac{f_{class,t}(t)A(t|nTP, TP_1, ..., TP_n)}{\Sigma_{n=1}^{nTP/2} \int_{TP_{2n-1}}^{TP_{2n}} f_{class,t}(t')\mathrm{d}t'}. \tag{4.24}$$

The denominator of this equation can be interpreted as the probability that the candidate should pass the selection criteria given its turning points. The normalisation for an exponential PDF convoluted with a Gaussian, equation 4.19, is given by [84]

$$\begin{aligned}\int_{t_1}^{t_2} \frac{1}{\tau} e^{-t_{meas}/\tau} e^{\frac{1}{2}\sigma_t^2/\tau^2} F\left(\frac{t}{\sigma_t} - \frac{\sigma_t}{\tau}\right) = {} & e^{-t_1/\tau} e^{\frac{1}{2}\sigma_t^2/\tau^2} F\left(\frac{t_1}{\sigma_t} - \frac{\sigma_t}{\tau}\right) - F\left(\frac{t_1}{\sigma_t}\right) \\ & - e^{-t_2/\tau} e^{\frac{1}{2}\sigma_t^2/\tau^2} F\left(\frac{t_2}{\sigma_t} - \frac{\sigma_t}{\tau}\right) + F\left(\frac{t_2}{\sigma_t}\right).\end{aligned} \tag{4.25}$$

Similarly the two exponential terms in the proper-decay-time PDF for secondary D^0, equation 4.21, yield normalisation terms exactly like this, multiplied by $\frac{1}{\tau}$. An example of a per-candidate acceptance function and the resulting proper-decay-time PDF for a prompt D^0 candidate is shown in figure 4.10.

Factorising out the PDF of the acceptance variables gives

$$\begin{aligned}f_{class,TPs}(nTP, TP_1, ..., TP_n) = {} & f_{class,nTP}(nTP) f_{class,TP1}(TP_1|nTP)... \\ & f_{class,TPn}(TP_n|nTP, TP_1, ..., TP_{n-1}).\end{aligned} \tag{4.26}$$

A PDF with so many correlated variables can be very difficult to work with. Exactly how to handle this depends on the distribution of turning points obtained for any given dataset. The specific implementation of this part of the PDF is thus discussed in section 5.3.1. In the simplest case this PDF may be the same for all classes of candidate in the dataset. In this case it factors out of the full fit PDF, equation 4.17, and has no effect on the determination of the optimal fit parameter values. It can thus be ignored in the fit PDF. However, should $f_{class,TPs}(nTP, TP_1, ..., TP_n)$ be different for different classes of candidate it must be included in the fit PDF; otherwise, the determination of the fit parameters will be biased as a result of the Punzi effect, as discussed in section 4.1.2.

4.3.2.3 Calculating the Average Proper-Decay-Time Acceptance

The full fit, performed using the per-candidate acceptance functions determined by the swimming algorithm, fits to the distributions of a large number of variables: t, $\ln(\chi^2(IP_{\mathrm{D}^0}))$, nTP, TP_1, ... , TP_n. Indeed, as nTP varies, the dimensionality of the fit is not constant - the fit can be viewed as several simultaneous fits to disjoint datasets, each with a different nTP. Visualising the results of such a fit is thus impossible without integrating out some of the fit variables. The most important visualisation to obtain is that of the projection of the fit onto the axis of the proper decay time, so that it can be overlaid on the proper-decay-time distribution of the data. This involves integrating out the dependence of the fit PDF on the turning point variables, and thus obtaining the average acceptance as a function of proper decay time for the data.

As was described in section 4.3.2.2 the PDF for each class for the lifetime fit takes the form

$$\begin{aligned} f_{class,T}(t, nTP, TP_1, ..., TP_n) &= f_{class,t}(t|nTP, TP_1, ..., TP_n) f_{class,TPs}(nTP, TP_1, ..., TP_n) \\ &= \frac{f_{class,t}(t)\left(\Sigma_{n=1}^{nTP/2}\Theta(t - TP_{2n-1})\Theta(TP_{2n} - t)\right)}{\Sigma_{n=1}^{nTP/2}\int_{TP_{2n-1}}^{TP_{2n}} f_{class,t}(t')\mathrm{d}t'} \\ & f_{class,TPs}(nTP, TP_1, ..., TP_n). \end{aligned} \tag{4.27}$$

To obtain the average proper-decay-time PDF for that class one then needs to integrate out

the acceptance variables

$$\begin{aligned} f_{class,t,avg}(t) &= \int_{-\infty}^{+\infty} f_{class,T}(t, nTP, TP_1, ..., TP_n)\mathrm{d}nTP\mathrm{d}TP_1...\mathrm{d}TP_n \\ &= f_{class,t}(t) \\ &\quad \int_{-\infty}^{+\infty} \left(\frac{\left(\Sigma_{n=1}^{nTP/2}\Theta(t - TP_{2n-1})\Theta(TP_{2n} - t)\right)}{\Sigma_{n=1}^{nTP/2}\int_{TP_{2n-1}}^{TP_{2n}} f_{class,t}(t')\mathrm{d}t'} \right. \\ &\qquad \left. f_{class,TPs}(nTP, TP_1, ..., TP_n) \right) \mathrm{d}nTP\mathrm{d}TP_1...\mathrm{d}TP_n \\ &\equiv f_{class,t}(t)A_{class,avg}(t). \end{aligned} \tag{4.28}$$

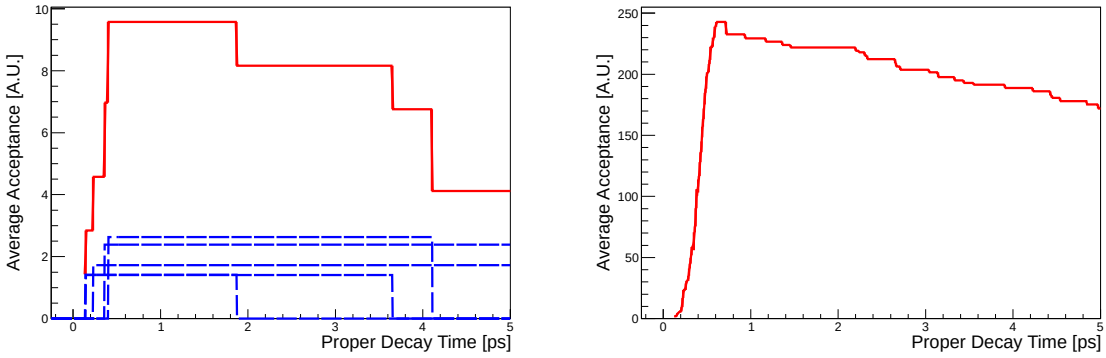
Thus the problem becomes one of calculating the average acceptance rate as a function of proper decay time, $A_{class,avg}(t)$, for each class. In the simple case that the distribution of turning points is the same for all classes the PDF $f_{class,TPs}(nTP, TP_1, ..., TP_n)$ can simply be modelled as a sum over each candidate in the dataset of Dirac δ-functions, taking the values of the turning points for each candidate. The integral over the turning points then simply becomes a sum, and the average acceptance function the sum of the acceptance functions for each candidate:

$$A_{class,avg}(t) = \Sigma_{i=0}^{N} \frac{\Sigma_{n=1}^{nTP_i/2}\Theta(t - TP_{i,2n-1})\Theta(TP_{i,2n} - t)}{\Sigma_{n=1}^{nTP_i/2}\int_{TP_{i,2n-1}}^{TP_{i,2n}} f_{class,t}(t')\mathrm{d}t'}. \tag{4.29}$$

Figure 4.11 shows examples of acceptance functions calculated in this manner. Here one sees that the contribution of each candidate increases as the probability of its being selected decreases, and that the acceptance function grows more smooth as the number of candidates increases. The existence of upper turning points also clearly decreases the average acceptance at higher proper decay times.

Note that, although the distribution of turning points may be the same for all classes of candidate, this does not mean that the average acceptance functions are the same. This is due to the weights in the denominator of equation 4.29. As the unbiased proper-decay-time PDF is different for each class so is this weight, and thus the resulting acceptance function differs as well. The case in which the distribution of turning points is different for each class is rather more complicated, and depends on the form of the PDF of the turning points. It is thus discussed in section 5.3.1 together with the specific $f_{class,TPs}(nTP, TP_1, ..., TP_n)$ used to perform the fits in chapter 5.

Having obtained the average acceptance functions for each class one can then compare the average proper-decay-time PDF obtained to the distribution of the data, and use this to calculate the χ^2 of the lifetime fit. Although this provides a much more comprehensible



(a) **(b)**

Figure 4.11: Examples of average acceptance functions calculated from toy data using equation 4.29. In (a) five candidates are used and the contribution of each candidate is shown by blue dashed lines, while the total average acceptance shown in red. (b) shows the average acceptance calculated using 100 candidates.

visualisation of the fit results it is very much a simplification of the full fit process. Thus one must also be careful to ensure that the PDFs of the turning point values used in the fit (assuming they need be included) describe the data well across the whole phase space of the fit.

4.4 The Full Fit PDF

With the PDFs determined as described in the previous sections of this chapter, one can express the full fit PDF as

$$\begin{aligned} f(m(\mathrm{D}^0),&\Delta m, t, \ln(\chi^2(IP_{\mathrm{D}^0})), nTP, TP_1, ..., TP_n) \\ &= \Sigma_{class}\big[f_{class,m}(m(\mathrm{D}^0))f_{class,\Delta m}(\Delta m) \\ &\quad f_{class,t}(t|nTP, TP_1, ..., TP_n)f_{class,IP}(\ln(\chi^2(IP_{\mathrm{D}^0}))|t)f_{class,TPs}(nTP, TP_1, ..., TP_n)\big]. \end{aligned} \tag{4.30}$$

Thus, one can perform the fit to the mass distributions first, to provide discrimination of any backgrounds that can be distinguished by their mass. This having been done, one can perform the fit to the proper-decay-time variables. At this stage the per-candidate acceptance variables are used to account for the selection bias, while the $\chi^2(IP_{\mathrm{D}^0})$ is used to distinguish secondary D^0. Thus full discrimination is achieved between the signal D^0 and all backgrounds, the selection bias is corrected, and one obtains the effective lifetime of the signal.

4.5 Possible Extensions

4.5.1 Accounting for Combinatorial and Three Body Backgrounds

As has been mentioned previously, combinatorial backgrounds are neglected in the fits performed to obtain the measurements presented in chapters 5 and 6. When the much larger dataset collected during 2011 is analysed such backgrounds must be taken into account, in order to achieve a sufficiently small systematic uncertainty on the measurements obtained. Also, when fitting for the $\mathrm{K^+K^-}$ lifetime one must account for three body backgrounds such as $\mathrm{D^0} \to \mathrm{K^-\pi^+\pi^0}$. Thus, these classes of background must be added to the fit, and their relevant PDFs obtained.

The PDFs for the mass distributions from such backgrounds should be well enough described by simple analytical functions. Obtaining accurate proper-decay-time PDFs for these backgrounds is rather more difficult. In [84, 85] a method is used whereby one uses the mass fit to calculate $P(\mathrm{background}|m)$. Then kernel density estimation is used, weighting each candidate's contribution by $P(\mathrm{background}|m)^2$, to obtain a proper-decay-time distribution that is dominated by the background. Using the known proper-decay-time PDF for the signal one can then subtract the remaining contribution of the signal to this distribution to obtain a PDF purely for background. There are a few caveats to this method that complicate its implementation. One of these is that it can only be used in the case that the proper-decay-time PDF is unknown for only one class of background. It has, nonetheless, proven effective under these circumstances.

A more general method is presented in the next section.

4.5.1.1 sWeights and sPlots

An elegant solution to the problem of multiple backgrounds with unknown proper-decay-time PDFs is to use 'sPlots', which are described in detail in [96]. This can to some extent be viewed as a more general extension of the previously described technique. It similarly uses a fit to some discriminating variables $\underline{\mathbf{X}}$, such as $m(\mathrm{D^0})$ and Δm, to calculate the probability of each candidate belonging to each class, $P(class|\underline{\mathbf{X}})$. From these a symmetric matrix is calculated which has the same dimension as the number of classes of candidates, and is defined by

$$\mathbf{V}^{-1}_{nm} = \frac{1}{N^2 P(class_n) P(class_m)} \Sigma_{i=1}^{N} P(class_n|\underline{\mathbf{X}}_i) P(class_m|\underline{\mathbf{X}}_i), \tag{4.31}$$

where the sum i is over all N candidates in the dataset, and the indices n and m represent the different candidate classes. From this one can define an 'sWeight' for each candidate for

each class

$$ {}_s\mathcal{P}_{class}(\underline{\mathbf{X}}_i) = \frac{\Sigma_{j=1}^{\text{N. classes}} \mathbf{V}_{class,j} f_j(\underline{\mathbf{X}}_i)}{N \Sigma_{k=1}^{\text{N. classes}} P(class_k) f_k(\underline{\mathbf{X}}_i)}, \tag{4.32} $$

where $f_j(\underline{\mathbf{X}})$ is the PDF of the variables $\underline{\mathbf{X}}$ for class j. If one then makes a histogram of a variable that is independent of those in $\underline{\mathbf{X}}$ and weights each candidates contribution by its sWeight for a specific class, one obtains the distribution of that variable for that specific class - the 'sPlot' of this variable. This is because the sWeights are calculated in such a way that the contributions from any other classes of candidate in the dataset cancel in the histogram. Thus, one can obtain separate distributions of a given variable for each class in the dataset, regardless of how many classes there are, and without needing to know the distribution for any classes beforehand. One caveat here is that the variable being plotted must be independent of the discriminating variables $\underline{\mathbf{X}}$.

Thus, using the mass as the discriminating variable, one can obtain the proper-decay-time distributions for all backgrounds, without needing to use the PDF for signal. Figure 4.12a shows a fit to the mass distribution of toy data with a signal and background class using 10,000 candidates of which 80 % are signal. The symmetric matrix used to calculate the sWeights in this case is

$$ \mathbf{V}_{nm} = \begin{pmatrix} 9512.2 & -1550.1 \\ -1550.1 & 3588.0 \end{pmatrix}. \tag{4.33} $$

There are various normalisation rules associated with sWeights. One can see that the sum of the elements in the first row is equal to the number of signal candidates (within statistical uncertainties), and the sum across the second row is equal to the number of background candidates. The sum of sWeights for a given candidate over all classes of candidate must also be equal to one. For example, using equation 4.32, the sWeights for a candidate with mass 1816.1 MeV, which has a high probability of being background, are

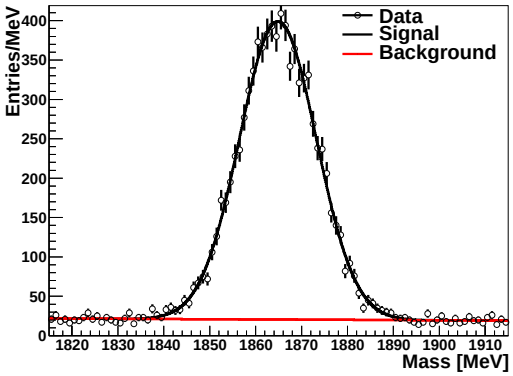
$$ {}_s\mathcal{P}_{signal}(1816.1\,\mathrm{MeV}) = \frac{9512.2 \times 4.1 \times 10^{-9} - 1550.1 \times 0.0104}{10,000 \times (0.8 \times 4.1 \times 10^{-9} + 0.2 \times 0.0104)} = -0.76 \tag{4.34a} $$

$$ {}_s\mathcal{P}_{bkg.}(1816.1\,\mathrm{MeV}) = \frac{-1550.1 \times 4.1 \times 10^{-9} + 3588.0 \times 0.0104}{10,000 \times (0.8 \times 4.1 \times 10^{-9} + 0.2 \times 0.0104)} = 1.76, \tag{4.34b} $$

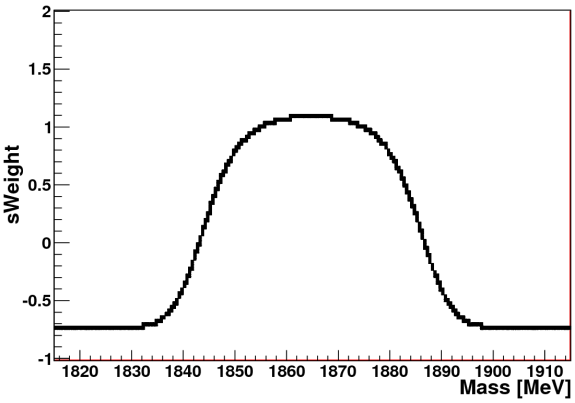
and for a candidate with mass 1863.5 MeV, which has a high probability of being signal, the sWeights are

$$ {}_s\mathcal{P}_{signal}(1863.5\,\mathrm{MeV}) = \frac{9512.2 \times 0.046 - 1550.1 \times 0.010}{10,000 \times (0.8 \times 0.046 + 0.2 \times 0.010)} = 1.09, \tag{4.35a} $$

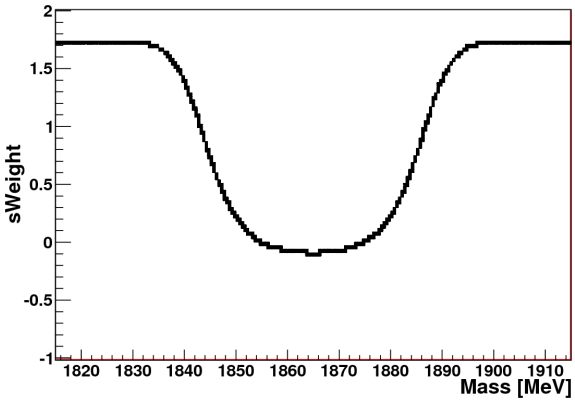
$$ {}_s\mathcal{P}_{bkg.}(1863.5\,\mathrm{MeV}) = \frac{-1550.1 \times 0.046 + 3588.0 \times 0.010}{10,000 \times (0.8 \times 0.046 + 0.2 \times 0.010)} = -0.09. \tag{4.35b} $$



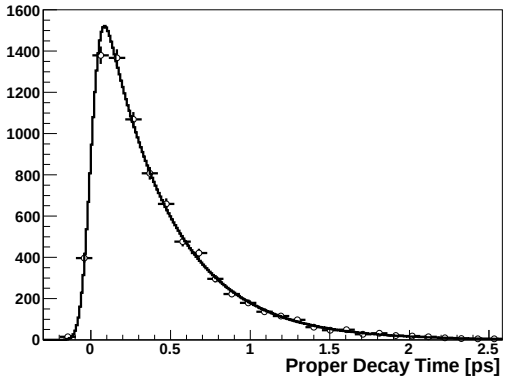
(a) The results of a fit to the mass distribution of simple toy data containing signal and background.



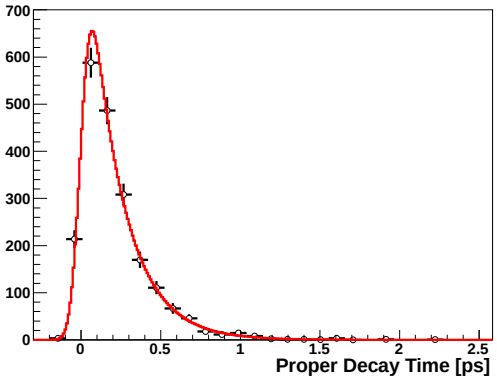
(b) The sWeight for signal as a function of mass, as calculated by equation 4.32.



(c) The sWeight for background as a function of mass, as calculated by equation 4.32.



(d) An sPlot of the proper decay time for signal, overlaid with the PDF used to generate the signal proper decay time.



(e) An sPlot of the proper decay time for background, overlaid with the PDF used to generate the background proper decay time.

Figure 4.12

Figures 4.12b and 4.12c shows the sWeights as a function of mass calculated using this fit for signal and background respectively. One can see that the sWeight for signal is positive in regions where a candidate has a high probability of being signal, and negative in regions where the the probability of being background is high. The opposite is seen in the sWeights for background. sPlots of the proper decay time for signal and background are shown in figures 4.12d and 4.12e respectively. These are overlaid with the PDFs used to generate the proper decay time for signal and background. There is an excellent agreement between the sPlots and the generation PDFs for both classes.

sWeights can also be used for kernel density estimation, so that one can obtain a continuous distribution. One issue in doing so is that sWeights can take negative values, and can thus give negative regions in the distribution obtained, particularly in regions of low density. This makes such a distribution unsuitable for use as a PDF. Simply setting these negative regions to zero is unsatisfactory as it results in discontinuities in the PDF. The technique described in [84] encounters a similar issue when the signal contribution is subtracted from the proper-decay-time distribution obtained by weighting by $P(\mathrm{background}|m)^2$. As this distribution has a certain amount of statistical fluctuation negative regions can result after the subtraction. The solution proposed in [84] is to compare the distribution prior to the subtraction, $f_0(t)$, to that after, $f_1(t)$. Defining the value

$$\epsilon(t) = \frac{f_1(t)}{f_0(t)}, \tag{4.36}$$

one can then map the interval $(-\infty, \epsilon_0) \to (0, \epsilon_0)$ using a function

$$\epsilon_c(t) = \begin{cases} \epsilon(t) & \text{if } \epsilon(t) \geq \epsilon_0, \\ \epsilon_0 \left(-\frac{2}{\pi} \arctan\left(\frac{\epsilon_0 - \epsilon(t)}{\epsilon_0} \right) + 1 \right) & \text{if } \epsilon(t) < \epsilon_0, \end{cases} \tag{4.37}$$

such that $\epsilon_c(t)$ is always positive. Here $\arctan$ is used due to its asymptotic behaviour, and could be replaced by any asymptotic function. This definition has the advantage of having a continuous first derivative. Applying the correction $f_1(t) = \epsilon_c(t) f_0(t)$ thus ensures that $f_1(t)$ is also always positive, and can thus be used as a PDF. The value of ϵ_0 varies the scale of the correction applied, and so must be tuned such that the distribution is not altered too much or too little. In [84] a value of $\epsilon_0 = 0.01$ is suggested. This method is used together with the method of obtaining a proper-decay-time PDF for combinatorial background described in [84] in the determination of the effective lifetime of $\mathrm{B}_s^0 \to \mathrm{K}^+\mathrm{K}^-$ detailed in [85].

The same method can be used to correct for negative regions when using sWeights in kernel density estimation. Having made the sPlot of a given variable for a given class one can also make the plot of the same variable and instead weight each candidate by $P(class|\underline{\mathbf{X}})^2$. One can then use this plot to apply exactly the same correction to the sPlot, and remove any

negative values. Thus, sPlots made using kernel density estimation can be used as PDFs in a fit.

4.5.1.2 Using sPlots to Obtain an Unbiased Proper-Decay-Time PDF

Making an sPlot of the proper-decay-time distribution from the data yields only the average distribution, with the acceptance effects averaged as well, as in equation 4.28. In order to use the per-candidate acceptance functions determined by the swimming algorithm one must have the unbiased proper-decay-time PDF for each class. Using equation 4.28 this can be obtained by dividing the average proper-decay-time distribution by the average acceptance as a function of proper decay time. The average acceptance can be calculated using equation 4.29, or the methods used for more complicated turning point distributions described in section 5.3.1.4. However, equation 4.29 also requires the unbiased proper-decay-time PDF. This issue can be circumvented by using an initial estimate of the unbiased proper-decay-time PDF when calculating the average acceptance. One can then divide the average proper-decay-time distribution by the acceptance function thus obtained to give the unbiased proper-decay-time PDF. Comparing this to the estimate PDF used in calculating the average acceptance one can check the accuracy of the initial estimate PDF, and thus of the unbiased PDF obtained. If they are different one can use the newly obtained unbiased PDF to recalculate the average acceptance. This process can be iterated upon until the unbiased PDF used to calculate the average acceptance, and the PDF obtained by dividing the average proper-decay-time distribution by the average acceptance agree sufficiently well.

Such an iterative process can have convergence issues in regions of very low acceptance. Thus, if it is necessary to use this process it may also be necessary to exclude the regions of lowest acceptance from the fit.

4.5.1.3 The sFit Method

Although using sPlots presents one method of dealing with additional backgrounds there is in fact a much simpler method of using sWeights to account for these backgrounds in the lifetime fit. This is the ‘sFit’ technique [89]. In this case the sWeights are calculated similarly, by using a fit to some discriminating variables, such as the mass. Then, rather than using the sWeights to make sPlots of the desired variables for the desired classes, each candidate’s contribution to the log likelihood of the proper-decay-time fit is instead weighted by its sWeight for the signal class. Just as in an sPlot, this use of sWeights means that the contribution of backgrounds cancels out of the total log likelihood of the data. Thus, only the signal component of the data contributes to the proper-decay-time fit. This means that

only the signal PDFs need be considered in the full proper-decay-time fit PDF. One need not know the proper-decay-time PDFs for any backgrounds that can be discriminated by the mass distribution. In the case of the D^0 this means that secondary D^0 would still need to be considered, as they cannot be discriminated by any variable that is not dependent on proper decay time. However, provided the PDFs used in the fit to the mass distributions are accurate, combinatorial backgrounds, and any other backgrounds such as $\mathrm{D}^0 \to \mathrm{K}^-\pi^+\pi^0$, can be cancelled out of the lifetime fit, and need not be parametrised.

One minor drawback of this method is that the statistical uncertainties, as estimated by the $\Delta\ln(L) = \frac{1}{2}$ technique used by MINUIT, can be underestimated. This is because the dataset is somewhat reduced in the cancellation of the backgrounds, reducing its statistical power. The level of this reduction is determined by how cleanly the signal and backgrounds are separated by the variables used to calculate the sWeights.

4.5.2 Obtaining an Accurate Parametrisation of Secondary D^0

Using the sFit technique, described in the previous section, would leave secondary D^0 as the only background to consider in the lifetime fit. Secondary D^0 are the primary source of background to $\mathrm{D}^0 \to hh'$ decays anyway, so improving their parametrisation in the fit is key to achieving high precision measurements. As discussed in section 4.2.2 parametric PDFs with sufficient flexibility as to allow them to fit the data are used in the measurements presented in chapters 5 and 6. However, the shape of such PDFs is still quite restricted, and so using non-parametric PDFs, such as those obtained using kernel density estimation, would be preferable. During the 2011 run large quantities of $\mathrm{B} \to \mathrm{D}^0(hh')X$ decays were collected. These candidates could potentially be reconstructed as prompt $\mathrm{D}^0 \to hh'$ and thus be used to obtain PDFs for secondary D^0. However, a careful study of trigger selection biases would be required.

4.6 Summary

This chapter presented in detail the methods used in performing fits to extract the effective lifetime of the D^0 meson in order to measure the parameters y_{CP} and A_Γ described in section 1.3. First, the general principles of fitting distributions of variables from data were introduced in section 4.1. This covered how the optimal values of the parameters of probability density functions (PDFs) can be found from the data, as well as the general form of PDFs used to describe datasets containing several classes of candidate, *i.e.* signal and backgrounds. The technique of obtaining continuous PDFs from data using kernel density estimation was also

introduced.

Section 4.2.1 then covered how to distinguish the majority of backgrounds using fits to mass distributions, and section 4.2.2 dealt specifically with how to distinguish secondary D^0, produced in B decays, from prompt signal. These techniques are used as described, where possible, in distinguishing signal and backgrounds in the fits performed to extract y_{CP} and A_Γ detailed in chapters 5 and 6.

The fit PDFs required to obtain the effective lifetime of the signal and to parametrise the proper-decay-time distribution of secondary D^0 were discussed in section 4.3.1. Section 4.3.2 described how to correct for lifetime biasing selection criteria in the trigger and offline candidate selections using the 'swimming algorithm'. Also covered was how to incorporate the acceptance variables determined by the swimming algorithm into the fit PDFs. This is used in chapters 5 and 6 to extract the average lifetimes of the D^0 and $\overline{\mathrm{D}}^0$ in order to calculate y_{CP} and A_Γ. The specifics of the PDF of the acceptance variables themselves are discussed in section 5.3.1, where kernel density estimation is used to obtain PDFs of the turning points from the data. The full fit PDF required to extract the effective lifetime of the signal was then presented in section 4.4. This is essentially that which is used to measure the lifetimes of D^0 and $\overline{\mathrm{D}}^0$ in chapter 5, with the exception that the level of combinatorial backgrounds is sufficiently low as to be neglected in the fit, as discussed in section 5.2. A corresponding systematic uncertainty for this simplification is assigned in section 6.3.3. Finally, possible improvements to the fit method, particularly involving the implementation of 'sWeights', were discussed in section 4.5.

Chapter 5

Measurement of the Charm Sector Mixing and CP-Violation Parameters y_{CP} and A_Γ

The fit methodology detailed in chapter 4 allows the extraction of the effective lifetime of a sample of D^0 candidates. These can then be used to calculate the parameters y_{CP} and A_Γ, as discussed in section 1.3 and defined in equations 1.71 and 1.74. Section 5.1 details the data used to do so, and the selection criteria applied to obtain the final datasets. Section 5.2 then describes the specific details and results of the fits performed to extract the signal yield. The specific methodology and results of the fits used to extract the effective lifetimes are presented in section 5.3. Finally, the resulting values of y_{CP} and A_Γ and their statistical uncertainties are shown. The stability of these measurements and their systematic uncertainties are evaluated in chapter 6.

5.1 Data Sample and Selection

The data used for the measurements presented here were collected by LHCb during September and October 2010 at $\sqrt{s} = 7\ \mathrm{TeV}$. The full sample comprises $28.0 \pm 2.8\,\mathrm{pb}^{-1}$. A further $\sim 10\,\mathrm{pb}^{-1}$ were collected prior to September 2010 but the swimming algorithm, detailed in section 4.3.2, was not commissioned for this period. Consequently it can only be applied to data collected after the release of v10r2 of the HLT software package MOORE.

There are several issues in triggering D^0 decays. Due to the very high production rate of D^0 at LHCb a clean signal sample can readily be obtained. The main issue is doing so while keeping the acceptance rate within the required limits. This rate must fit within the timing constraints of the HLT, and stay within the limits of the available permanent storage space.

The configuration of the L0 trigger varied somewhat during the relevant running period, to account for changes in running conditions. As described in section 2.2.7 it requires a high p_T track segment in the muon chambers or high E_T cluster in the calorimeters and also applies a maximum limit on the particle multiplicity, as estimated by the SPD and VELO VETO stations. The value of the E_T cut greatly influences the number of D^0 decays selected. Although the tracks made by the daughters of a D^0 tend to be among the highest E_T tracks in an event, their E_T is still rather low compared to the daughters of a B. Thus, too tight an E_T cut can result in a very low selection efficiency for D^0 decays. Consequently the value of the E_T cut at L0 requires careful tuning to allow a reasonable sample of D^0 to be collected while keeping the L0 acceptance rate below the maximum input rate to the HLT. During the relevant running period the cut value varied between $E_T > 2.26\,\mathrm{GeV}$ and $E_T > 4.14\,\mathrm{GeV}$. The majority of the data were taken using $E_T > 3.6\,\mathrm{GeV}$.

The specific HLT lines relevant to the data studied here are the HLT1 '1 Track' and HLT2 'D2hh' triggers. As its name suggests, the '1 Track' trigger requires only one track with a large impact parameter to trigger an event. The 'D2hh' trigger fully reconstructs D^0 candidates under four mass hypotheses: $\mathrm{D}^{*\pm} \to \overset{(-)}{\mathrm{D}^0}\pi_s^\pm$ with $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$, $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$, $\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^+\mathrm{K}^-$ and $\overset{(-)}{\mathrm{D}^0} \to \pi^+\pi^-$. Events with candidates passing the selection under any of these hypotheses are triggered. The HLT faces similar issues as L0 in reducing the event rate to that allowed by the capacity of the permanent storage. Hence, some very tight lifetime biasing selection criteria are used. The cuts used during the relevant data-taking period are listed in table 5.1. To further reduce the retention rate the candidates in HLT2 are required to have reconstructed mass within 25 MeV of the D^0 mass. This requirement removes much of the mass side-bands that contain primarily combinatorial backgrounds. To ensure that some candidates from the mass side-bands are retained an almost identical HLT2 trigger line is run with a much wider mass window of $\pm$ 250 MeV about the D^0 peak - the 'D2hh Wide Mass' trigger. To keep the acceptance rate low a 'prescale' of exactly 1 % is applied to this trigger, meaning that a random sample consisting of only 1 % of the events passing this trigger are kept. Though some cut values differ between the 'D2hh' and 'D2hh Wide Mass' triggers those that affect the reconstructed mass distributions are superseded by tighter cuts in the offline selection; those uncorrelated to the reconstructed mass only change the relative retention rate between the two trigger lines slightly.

An 'untagged' trigger line, where the $\mathrm{D}^{*\pm}$ is not reconstructed and so the flavour of the D^0 is unknown, is also run. During the 2010 run this trigger line applied a tight cut of $\chi^2(IP_{\mathrm{D}^0}) < 2$. While this requirement removes most of the contribution from secondary D^0 there remains a small, irreducible fraction in the dataset. As secondary D^0 can only be cleanly distinguished from prompt D^0 at high $\chi^2(IP_{\mathrm{D}^0})$ this cut makes it exceptionally

Trigger	Selection
HLT1 '1 Track'	Single track $\|IP\| > 110\,\mu\text{m}$
	Single track $\chi^2_{IP} > 34$
	Single track N. VELO Hits > 9
HLT2 'D2hh'	Daughter $\chi^2_{IP} > 2$
	D^0 $\chi^2(\text{flight distance}) > 25$
	D^0 $\cos(\text{pointing angle}) > 0.99985$
	$1864.8 - 25 < m(\text{D}^0)[\,\text{MeV}] < 1864.8 + 25$

Table 5.1: The salient trigger criteria applied to the data studied in this chapter, using the HLT1 '1 Track' and HLT2 'D2hh' triggers.

difficult to achieve any discrimination of secondary D^0. Thus, the dataset collected using the untagged trigger line in 2010 cannot be used in measurements of the effective lifetime of the D^0. For this reason only the dataset collected using the tagged trigger line is used in the measurements presented here.

The HLT selections produce a very clean sample of D^0 candidates with a high signal fraction. Thus the stripping selection serves only to ensure that the selected candidates are of good quality, once the full offline reconstruction is run. The selection criteria applied are shown in table 5.2. Here the datasets for each mass hypothesis are made disjoint by the use of delta-log-likelihood (DLL) cuts on the PID. The likelihood of a given track being of a given species is calculated using the information provided by the RICH detectors, as described in section 2.2.2.2. $\text{DLL}_{\text{K}\pi}$ is the difference in the natural logarithm of the likelihood of the given track corresponding to a K and a π. A positive $\text{DLL}_{\text{K}\pi}$ thus implies that the track is more likely to have been made by a K than a π, while a negative $\text{DLL}_{\text{K}\pi}$ implies the opposite. Requiring a positive $\text{DLL}_{\text{K}\pi}$ for K candidates and a negative $\text{DLL}_{\text{K}\pi}$ for π candidates thus ensures that $\text{D}^0 \to \text{K}^-\pi^+$, $\overline{\text{D}}{}^0 \to \text{K}^+\pi^-$, $\overset{(-)}{\text{D}}{}^0 \to \text{K}^+\text{K}^-$ and $\overset{(-)}{\text{D}}{}^0 \to \pi^+\pi^-$ candidates all form disjoint datasets. As the majority of tracks are made by πs the probability of mis-identifying a π as a K is higher than that of mis-identifying a K as a π. Hence, a tighter DLL cut is used to identify Ks than is used to identify πs.

The fit range is restricted to specific intervals in the reconstructed mass of the D^0, $m(\text{D}^0)$, the difference between $m(\text{D}^0)$ and the reconstructed mass of the D^{*+}, Δm, and the proper decay time of the D^0, t. Thus, the final datasets used are those defined to lie in the region $1864 - 16 < m(\text{D}^0)\,[\,\text{MeV}] < 1864 + 16$, $145.4 - 2 < \Delta m\,[\,\text{MeV}] < 145.4 + 2$, and $0.25 < t\,[\,\text{ps}] < 6.0$. To obtain access to the mass side-bands candidates with $m(\text{D}^0)$ in the regions $[1815, 1848]$ MeV and $[1880, 1915]$ MeV obtained using the 'D2hh Wide Mass' trigger are

Type	Value
Daughters	$p > 5.0$ GeV
	$p_T > 0.9$ GeV
	Track χ^2/NDOF < 5
	K $\mathrm{DLL_{K\pi}} > 8$
	π $\mathrm{DLL_{K\pi}} < -5$
$\mathrm{D^0}$	Vertex χ^2/NDOF < 13
	$p_T > 3.3$ GeV
	Flight distance > 0.9 mm
$\mathrm{D^{*+}}$	Vertex χ^2/NDOF < 13
	$p_T > 3.6$ GeV
	π_s $p_T > 260$ MeV

Table 5.2: The stripping selection criteria applied after offline reconstruction. Here $\mathrm{DLL_{K\pi}}$ is the difference between the natural logarithm of the likelihood of the given track corresponding to a K and that of it corresponding to a π (the 'delta-log-likelihood' between the K and π hypotheses).

included in the datasets. The acceptance interval in proper decay time is applied to remove the region of very low statistics at low proper decay times, which can cause instabilities in the fit, and to exclude very long lived backgrounds, which can bias the fit results. These cuts are included in the per-candidate acceptance functions as minimum and maximum turning points. Additionally, as was discussed in section 4.3.2.1, only events with a single candidate are used, and both tracks of the candidates selected are required to have fired the HLT1 '1 Track' trigger. Finally, all candidates are required to have at most six turning points in the acceptance function determined by the swimming algorithm, in order to limit the number of fit variables. This is a very loose selection criterion, as very few candidates have more than two turning points, as shown in figure 5.8a. The number of candidates fulfilling these criteria for each decay channel are shown in table 5.3. The untagged datasets, used to calculate y_{CP}, are simply the combined datasets of $\mathrm{D^0}$ and $\mathrm{\overline{D}^0}$ in each decay mode. The decay $\mathrm{D^0 \to K^-\pi^+}$ is Cabibbo favoured, while $\mathrm{D^0 \to K^+K^-}$ is Cabibbo suppressed, thus significantly more $\mathrm{D^0 \to K^-\pi^+}$ than $\mathrm{D^0 \to K^+K^-}$ candidates are reconstructed. As the decay $\mathrm{D^0 \to \pi^+\pi^-}$ is doubly Cabibbo suppressed the number of candidates selected in this channel is only only ∼40 % that of $\mathrm{D^0 \to K^+K^-}$. The inclusion of these candidates would require additional background studies and would yield only a ∼15 % improvement in precision on $\tau_{\mathrm{eff}}(\mathrm{\overset{(-)}{D}{}^0} \to h^+h^-)$. Thus, only $\mathrm{D^0 \to K^+K^-}$ candidates are currently used when

	D^0	$\overline{\mathrm{D}}^0$	Untagged
$\mathrm{K}^+\mathrm{K}^-$	19,717	19,546	39,263
$\mathrm{K}\pi$	140,814	145,345	286,159

Table 5.3: The number of candidates used in the fits to extract y_{CP} and A_Γ, including backgrounds.

measuring y_{CP} and A_Γ.

5.2 Extraction of the Signal Yields

5.2.1 The Specific Methods Used to Fit the Distributions of $m(\mathrm{D}^0)$ and Δm

The methods used to fit the distributions of $m(\mathrm{D}^0)$ and Δm, in order to extract the relative fractions of signal D^0, randomly-tagged D^0 and combinatorial backgrounds, are those described in section 4.2.1. Secondary D^0 cannot be distinguished using the mass distributions and so are included in the signal class at this stage. In order to increase the fit speed the fits to the distributions of $m(\mathrm{D}^0)$ and Δm are performed separately. The specific PDFs are chosen as those that best describe the data. The results of the fits using these PDFs to the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$ datasets are presented in section 5.2.2.

As discussed in section 4.2.1 the PDF of $m(\mathrm{D}^0)$ for signal candidates can be described by a sum of several Gaussian functions with a common mean. A sum of two Gaussians is found to describe the data sufficiently well, so that the PDF is given by

$$f_{signal,m}(m(\mathrm{D}^0)) = f_{m1}Gauss(m(\mathrm{D}^0), \mu_m, \sigma_{m1}) + (1 - f_{m1})Gauss(m(\mathrm{D}^0), \mu_m, \sigma_{m2}). \tag{5.1}$$

No sensitivity to a lower radiative tail in the signal mass distributions is observed in any of the decay channels. This may need to be considered in future when dealing with larger datasets and the PDF adjusted accordingly. For combinatorial backgrounds the mass distribution is modelled using a straight line, normalised within the fit range

$$f_{bkg,m}(m(\mathrm{D}^0)) = \begin{cases} \frac{1}{m_{max}-m_{min}} - \frac{1}{2}K_{bkg}(m_{max}+m_{min}) + K_{bkg}m(\mathrm{D}^0) & \text{if } m_{min} < m(\mathrm{D}^0) < m_{max}, \\ 0 & \text{if } m(\mathrm{D}^0) < m_{min} \text{ or } m_{max} < m(\mathrm{D}^0). \end{cases} \tag{5.2}$$

Parameter	Class	Description
f_{signal}	Both	The fraction of signal D^0 (including secondary and randomly-tagged D^0) in the dataset
μ_m	Signal	The mean mass of the signal
f_{m1}	Signal	The relative fraction of the first Gaussian
σ_{m1}	Signal	The σ of the first Gaussian
σ_{m2}	Signal	The σ of the second Gaussian
K_{bkg}	Combi. Background	The gradient of the linear mass distribution

Table 5.4: The parameters of the fit to $m(\mathrm{D}^0)$ and their meaning. The class of candidate to which they apply is also shown; this can be signal D^0 (including secondary and randomly-tagged D^0), combinatorial background, or both.

The full list of parameters in this fit is shown in table 5.4.

Due to the retention of only 1 % of candidates passing the 'D2hh Wide Mass' trigger there are very few candidates in the mass side-bands. A scale factor is introduced for the PDFs in the side-band regions to account for the difference in acceptance rate between the two triggers. This factor is calculated simply as the ratio of the number of candidates in the signal region obtained using the 'Wide Mass' trigger to the number in this region obtained using the signal trigger. This transpires to be very close to the 1 % relative retention rate. The fit to the mass distribution is then performed to obtain the fraction of combinatorial background in the datasets.

To determine the fraction of randomly-tagged D^0, as is required for correct measurement of A_Γ, a fit is then done to the distribution of Δm. This is performed using candidates in the range $140 < \Delta m\,[\mathrm{MeV}] < 160$, not just in the signal region of $145.4 - 2 < \Delta m\,[\mathrm{MeV}] < 145.4 + 2$. All other selection criteria are unchanged. In this fit the PDF for correctly tagged signal is modelled as the sum of two Gaussians with a common mean, plus a third with an independent mean

$$\begin{aligned} f_{signal,\Delta m} =& f_{\Delta m1} Gauss(\Delta m, \mu_{\Delta m1}, \sigma_{\Delta m1}) + f_{\Delta m2} Gauss(\Delta m, \mu_{\Delta m1}, \sigma_{\Delta m2}) \\ &+ (1 - f_{\Delta m1} - f_{\Delta m2}) Gauss(\Delta m, \mu_{\Delta m2}, \sigma_{\Delta m3}). \end{aligned} \tag{5.3}$$

The third Gaussian is added to provide a sufficiently good description of the data. Although its addition is not physically motivated, studies have shown that the candidates described by this Gaussian are indeed signal-like [97].

Combinatorial background and randomly-tagged D^0 are modelled as a single class of candidate with the same PDF for Δm. The 'RooDstD0Bg' PDF, defined in the ROOFIT

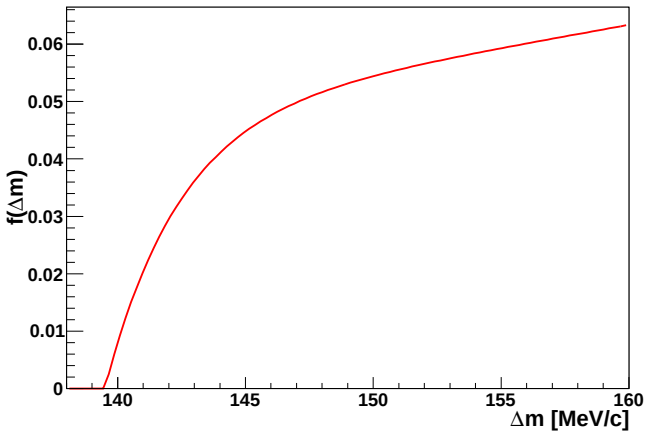


Figure 5.1: An example of the RooDstD0Bg PDF, as defined in equation 5.4, with $A = -0.34$, $B = 0.01$, $C = 2.8$ and $D = 139.5$.

package [98], is used to describe their distribution. It is defined as

$$f(\Delta m) = \begin{cases} \left(\frac{\Delta m}{D}\right)^A \left(1 - e^{-(\Delta m - D)/C}\right) + B\left(\frac{\Delta m}{D} - 1\right) & \text{if } \Delta m \geq D \\ 0 & \text{if } \Delta m < D. \end{cases} \tag{5.4}$$

The parameter D defines the turn-on point of the distribution, which should be equal to $m(\pi^+)$ as discussed in section 4.2.1; C determines the curvature of the distribution at low Δm; A and B define its slope at high Δm. An example of the shape of this PDF is shown in figure 5.1. The list of parameters in the Δm fit are given in table 5.5.

To aid convergence of the fit, the B parameter of this PDF is fixed to 0.01. The fraction of candidates determined to be of this class thus corresponds to the fraction of combinatorial background plus the fraction of randomly-tagged D^0. The fraction of correctly tagged D^0 lying within the signal region $145.4-2 < \Delta m\,[\mathrm{MeV}] < 145.4+2$ is then calculated to obtain the rate relevant to the final measurements. This is done by integrating the respective PDFs over the signal region and recalculating their relative fractions. The fraction of correctly tagged signal is then subtracted from the fraction of all signal determined by the fit to $m(\mathrm{D}^0)$ to obtain the random-tag rate, $f_{\textit{random-tag signal}} = f_{signal} - f_{\textit{tagged signal}}$. As discussed in section 4.2.1, as π^+ and π^- are produced in equal quantities, candidates will only be assigned the wrong flavour tag at half of this rate.

5.2.2 Measurements of Signal Yield and Random-Tag Rate

The number of candidates in the mass side-bands for the $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ datasets is so low that the fit to the distribution of $m(\mathrm{D}^0)$ cannot converge when a combinatorial background PDF is included. Thus, in order to evaluate the signal yield and random-tag rate only $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$

Parameter	Class	Description
$f_{tagged\,signal}$	Both	The fraction of signal that have the $\mathrm{D}^{*\pm}$ correctly reconstructed, and so have the correct flavour tag (including secondary D^0)
$\mu_{\Delta m1}$	Signal	The mean Δm for the first two Gaussians
$f_{\Delta m1}$	Signal	The relative fraction of the first Gaussian
$\sigma_{\Delta m1}$	Signal	The σ of the first Gaussian
$f_{\Delta m2}$	Signal	The relative fraction of the second Gaussian
$\sigma_{\Delta m2}$	Signal	The σ of the second Gaussian
$\mu_{\Delta m2}$	Signal	The mean of the third Gaussian
$\sigma_{\Delta m3}$	Signal	The σ of the third Gaussian
A_{bkg}	Background	Controls the slope of the PDF at high Δm
B_{bkg}	Background	Controls the slope of the PDF at high Δm - fixed to 0.01 in the fit
C_{bkg}	Background	Controls the slope of the PDF at low Δm
D_{bkg}	Background	The turn-on point of the PDF

Table 5.5: The parameters of the fit to Δm, and their meaning. The class to which they relate is also shown; this can be signal (including secondaryD^0), background (randomly-tagged D^0 and combinatorial background combined), or both.

candidates are used. This can be done as the random-tag rate is independent of the D^0 decay channel, and $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ provides significantly more candidates than $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$. The implications of this for the fit to the proper-decay-time distribution and the determination of the signal effective lifetimes are discussed in section 5.3.1.

The fits to $m(\mathrm{D}^0)$ and Δm for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$ are shown in figure 5.2. The pull of the fit in each bin is shown below each plot. The pull is calculated as the number of candidates minus the value of the fit PDF, divided by the statistical error on the number of candidates. The pulls are evenly distributed about zero, showing that the PDFs describe the data well. The corresponding fitted parameter values are shown in table 5.6. There is $\sim$99.2 % signal in the datasets used to determine the effective lifetimes, of which $\sim$95.7 % has the $\mathrm{D}^{*\pm}$ correctly reconstructed. Although the analogous fits to the $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ mass distributions do not converge stably they predict the fraction of combinatorial background to be $\sim$3 %.

The mean $m(\mathrm{D}^0)$ obtained is 1864.05 ± 0.18 MeV, which is slightly lower that the world average value of 1864.80 ± 0.14 MeV [7]. A low mean reconstructed mass is a common issue in the reconstruction at LHCb, as the momentum scale is slightly biased. Indeed, the fitted mean of Δm for signal is 145.4075 ± 0.0021 MeV, which is reasonably consistent

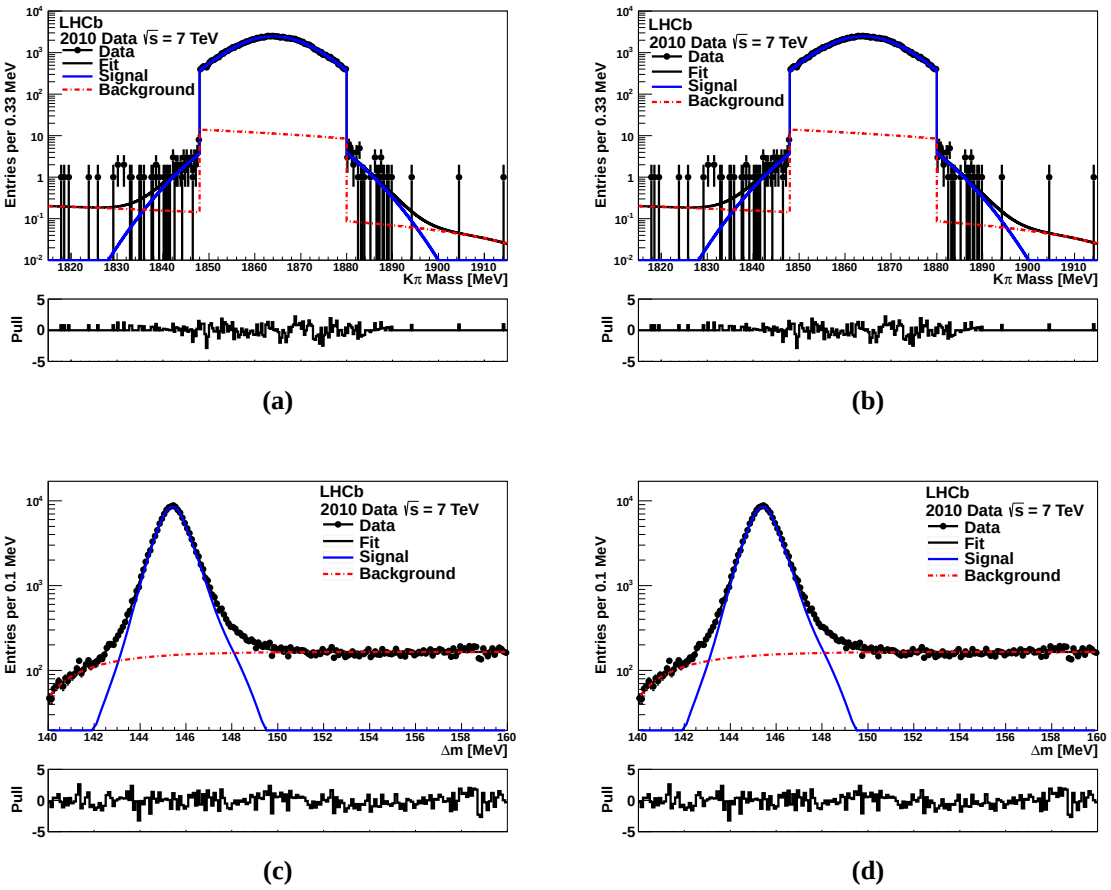


Figure 5.2: (a) and (b): The fits to the distributions of $m(\mathrm{D}^0)$ for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$ respectively, overlaid on the fitted PDFs for signal (including randomly-tagged D^0 and secondary D^0) and combinatorial background. (c) and (d): The fits to the distributions of Δm for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$ respectively, overlaid with the fitted PDFs for signal (including secondary D^0) and background (comprising combinatorial background and randomly-tagged D^0). The pull of the fit in each bin is shown below each plot. The corresponding fitted parameter values are shown in table 5.6.

$D^0 \to K^-\pi^+$ Mass Fit Results	
Parameter	Fitted Value
f_{signal}	0.9921 ± 0.0024
μ_m	1864.050 ± 0.026 MeV
f_{m1}	0.33 ± 0.15
σ_{m1}	6.17 ± 0.51 MeV
σ_{m2}	9.43 ± 0.82 MeV
K_{bkg}	$(-1.56 \pm 0.40) \times 10^{-4}\,\mathrm{MeV}^{-1}$

(a)

$\bar{D}^0 \to K^+\pi^-$ Mass Fit Results	
Parameter	Fitted Value
f_{signal}	0.9929 ± 0.0033
μ_m	1864.040 ± 0.026 MeV
f_{m1}	0.708 ± 0.044
σ_{m1}	7.06 ± 0.14 MeV
σ_{m2}	13.4 ± 1.0 MeV
K_{bkg}	$(-2.0 \pm 2.6) \times 10^{-4}\,\mathrm{MeV}^{-1}$

(b)

$D^0 \to K^-\pi^+$ Δm Fit Results	
Parameter	Fitted Value
$f_{tagged\,signal}$	0.8226 ± 0.0021
$\mu_{\Delta m1}$	145.4080 ± 0.0029 MeV
$f_{\Delta m1}$	0.192 ± 0.029
$\sigma_{\Delta m1}$	0.365 ± 0.019 MeV
$f_{\Delta m2}$	0.647 ± 0.034
$\sigma_{\Delta m2}$	0.715 ± 0.043 MeV
$\mu_{\Delta m2}$	145.731 ± 0.046 MeV
$\sigma_{\Delta m3}$	1.412 ± 0.057 MeV
A_{bkg}	-0.07 ± 0.52
B_{bkg}	0.01
C_{bkg}	2.59 ± 0.41
D_{bkg}	139.16 ± 0.18 MeV
$f_{tagged\,signal}$	0.9581 ± 0.0065

(c)

$\bar{D}^0 \to K^+\pi^-$ Δm Fit Results	
Parameter	Fitted Value
$f_{tagged\,signal}$	0.8208 ± 0.0020
$\mu_{\Delta m1}$	145.4070 ± 0.0030 MeV
$f_{\Delta m1}$	0.205 ± 0.038
$\sigma_{\Delta m1}$	0.374 ± 0.021 MeV
$f_{\Delta m2}$	0.590 ± 0.045
$\sigma_{\Delta m2}$	0.689 ± 0.045 MeV
$\mu_{\Delta m2}$	145.647 ± 0.032 MeV
$\sigma_{\Delta m3}$	1.284 ± 0.044 MeV
A_{bkg}	-0.64 ± 0.55
B_{bkg}	0.01
C_{bkg}	2.97 ± 0.55
D_{bkg}	138.92 ± 0.23 MeV
$f_{tagged\,signal}$	0.9569 ± 0.0068

(d)

Table 5.6: (a) The fitted values with their statistical uncertainties of the parameters of the fit to $m(D^0)$ using $D^0 \to K^-\pi^+$, and (b) the same for $\bar{D}^0 \to K^+\pi^-$. The parameter definitions are shown in table 5.4 on p. 125. (c) The fitted values and their statistical uncertainties of the fit to Δm using $D^0 \to K^-\pi^+$, and (d) the same for $\bar{D}^0 \to K^+\pi^-$. The final rows in (c) and (d) show the values of $f_{tagged\,signal}$ when calculated in the signal region, $145.4 \pm 2\,\mathrm{MeV}$. The parameter definitions are shown in table 5.5 on p. 127. The corresponding fitted distributions are shown in figure 5.2.

with the expected value of 145.421 $\pm$ 0.010 MeV [7], showing that $m(\mathrm{D}^0)$ and $m(\mathrm{D}^{*+})$ are biased by the same amount. This bias to the reconstructed $m(\mathrm{D}^0)$ is sufficiently small as not to be an issue in the determination of the effective lifetimes. The proper decay time of a given candidate is calculated as $t = m(\mathrm{D}^0)D_F/p$, where D_F is the distance of flight in the detector. As the reconstructed $m(\mathrm{D}^0)$ is determined by the momenta of its daughters it is biased by almost exactly the same factor as p. Thus, this bias cancels almost completely in the calculation of the proper decay time. Any remaining bias will cancel in the calculation of y_{CP} and A_Γ. The bias to the mean reconstructed $m(\mathrm{D}^0)$ can thus be safely neglected in the measurements presented here.

5.3 Extraction of the Effective Lifetimes

5.3.1 Specific Methodology of the Simultaneous Fit to Proper Decay Time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$

The method of performing the simultaneous fit to the distributions of proper decay time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ are for the most part those detailed in sections 4.2.2 and 4.3.2. However, as mentioned in section 5.2.2, so few candidates are accepted by the 'Wide Mass' trigger line in the mass side-bands that the fit to the distribution of $m(\mathrm{D}^0)$ cannot be used to distinguish combinatorial backgrounds for $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$. Similarly, very little statistical discrimination is achieved for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$. This means that none of the methods detailed in section 4.5 can be used to obtain a proper-decay-time PDF for combinatorial backgrounds. Attempting to use a parametric PDF, such as a sum of exponentials, also results in a fit that cannot converge stably. However, the level of combinatorial backgrounds in the signal region is $\sim$1 % for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$, as shown in table 5.6, and $\sim$3 % for $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$. Given the precision that can be achieved using the statistics given in table 5.3 it is reasonable to neglect backgrounds at this level in the fit to the proper-decay-time distribution without significantly degrading the precision of the result. Doing so means that the proper-decay-time component of the fit only considers prompt and secondary D^0. A systematic uncertainty resulting from this approximation is calculated in section 6.3.

As the mass parameters provide no discrimination between prompt and secondary D^0 these are omitted from the full fit PDF, which takes the form

$$f(\underline{\mathbf{T}}) = \Sigma_{class} f_{class,t}(t|nTP, TP_1, ..., TP_n) f_{class,IP}(\ln(\chi^2(IP_{\mathrm{D}^0}))|t) \\ f_{class,TPs}(nTP, TP_1, ..., TP_n) P(class). \quad (5.5)$$

The exact form of the PDF of the acceptance variables, $f_{class,TPs}(nTP, TP_1, ..., TP_n)$, is

discussed in section 5.3.1.3.

5.3.1.1 Accounting for Mis-Tagged D^0

The effects of mis-tagged D^0 can be accounted for after the signal lifetime has been obtained by using the random-tag rates determined by the fits to $m(\mathrm{D}^0)$ and Δm using the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ datasets. Treating mis-tagged D^0 as signal in the determination of the effective lifetime introduces a small bias, depending on the mis-tag rate. Assuming no production asymmetry for $\pi^\pm$ the fraction of candidates that have the wrong flavour tag will be half of the random-tag rate determined by the mass distribution fits. For a small mis-tag rate the measured effective lifetimes will be

$$\tau_{\mathrm{meas}}(\mathrm{D}^0) \simeq (1-\epsilon_+)\tau_{\mathrm{eff}}(\mathrm{D}^0) + \epsilon_+\tau_{\mathrm{eff}}(\overline{\mathrm{D}}^0), \tag{5.6a}$$

$$\tau_{\mathrm{meas}}(\overline{\mathrm{D}}^0) \simeq (1-\epsilon_-)\tau_{\mathrm{eff}}(\overline{\mathrm{D}}^0) + \epsilon_-\tau_{\mathrm{eff}}(\mathrm{D}^0), \tag{5.6b}$$

where ϵ_+ is the fraction of the candidates tagged as D^0 that are actually $\overline{\mathrm{D}}^0$, and vice versa for ϵ_-. These are found using $\epsilon_\pm = \frac{1}{2} f_{\textit{random-tag signal}}$ for the D^0 and $\overline{\mathrm{D}}^0$ datasets respectively. These can be solved to remove the bias from mis-tagged candidates to give

$$\tau_{\mathrm{eff}}(\mathrm{D}^0) \simeq \frac{1}{1-\epsilon_+-\epsilon_-}\left((1-\epsilon_-)\tau_{\mathrm{meas}}(\mathrm{D}^0) - \epsilon_+\tau_{\mathrm{meas}}(\overline{\mathrm{D}}^0)\right), \tag{5.7a}$$

$$\tau_{\mathrm{eff}}(\overline{\mathrm{D}}^0) \simeq \frac{1}{1-\epsilon_+-\epsilon_-}\left((1-\epsilon_+)\tau_{\mathrm{meas}}(\overline{\mathrm{D}}^0) - \epsilon_-\tau_{\mathrm{meas}}(\mathrm{D}^0)\right). \tag{5.7b}$$

A_Γ is then calculated using these corrected lifetimes. Using the results of the fits to $m(\mathrm{D}^0)$ and Δm using $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$, shown in table 5.6, yields

$$\epsilon_+ = \quad 0.017 \pm 0.003, \quad \text{and} \tag{5.8a}$$

$$\epsilon_- = \quad 0.018 \pm 0.003, \tag{5.8b}$$

where the uncertainties are the sum in quadrature of the statistical uncertainty and systematic uncertainties from the accuracy of integrating the PDFs, the accuracy of the fit model, and the $\mathrm{D}^{*\pm}$ production asymmetry. As y_{CP} uses the effective lifetimes of the untagged samples no correction for mis-tagged D^0 need be applied before calculating its value.

5.3.1.2 The Proper-Decay-Time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDFs

The chosen PDFs of proper decay time for prompt and secondary D^0 are those discussed and motivated in section 4.3.1. The distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ has no physically motivated shape, and so the PDFs of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ are chosen to give a sufficiently good description of the data. The fits made using these PDFs are presented in section 5.3.2.

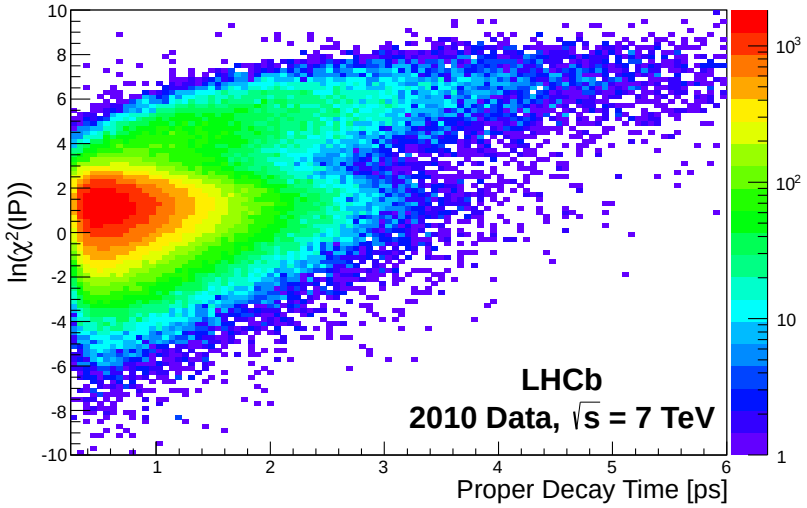


Figure 5.3: The distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time for the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ dataset.

As discussed in section 4.3.1, the proper-decay-time PDF for prompt D^0 is taken as the convolution of an exponential PDF with a Gaussian resolution PDF. For secondary D^0 a convolution of two exponential PDFs with a Gaussian is used. This approximation was shown to describe data obtained from full Monte Carlo simulation sufficiently well. One τ parameter of this PDF is fixed to 270 fs, to aid the fit convergence. This value is obtained from the fit to the sample of secondary D^0 obtained from Monte Carlo simulated data shown in figure 4.6. The σ of the Gaussian component that corresponds to the effect of proper-decay-time resolution is fixed to 50 fs for both PDFs. Under the assumption that the proper-decay-time measurements are unbiased, the mean of the Gaussian resolution function is fixed to zero.

Figure 5.3 shows the distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time, for the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ dataset. This clearly shows a dominant proper-decay-time independent component peaking around $\ln(\chi^2(IP_{\mathrm{D}^0})) = 1.5$, from prompt D^0; there is also clearly a small fraction of the data for which the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ increases with proper decay time, which corresponds to secondary D^0. The distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ roughly follows the shape of a bifurcated Gaussian - a Gaussian function with different σ values below and above the mean. This is defined as

$$Bif.Gauss(x,\mu,\sigma_{low},\sigma_{high}) = \begin{cases} \frac{2\sigma_{low}}{\sigma_{low}+\sigma_{high}} Gauss(x,\mu,\sigma_{low}) & \text{if } x \leq \mu, \\ \frac{2\sigma_{high}}{\sigma_{low}+\sigma_{high}} Gauss(x,\mu,\sigma_{high}) & \text{if } x > \mu. \end{cases} \tag{5.9}$$

Figure 5.4 shows the distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time for prompt D^0, taken from full Monte Carlo simulated data. Also shown are the values of μ,

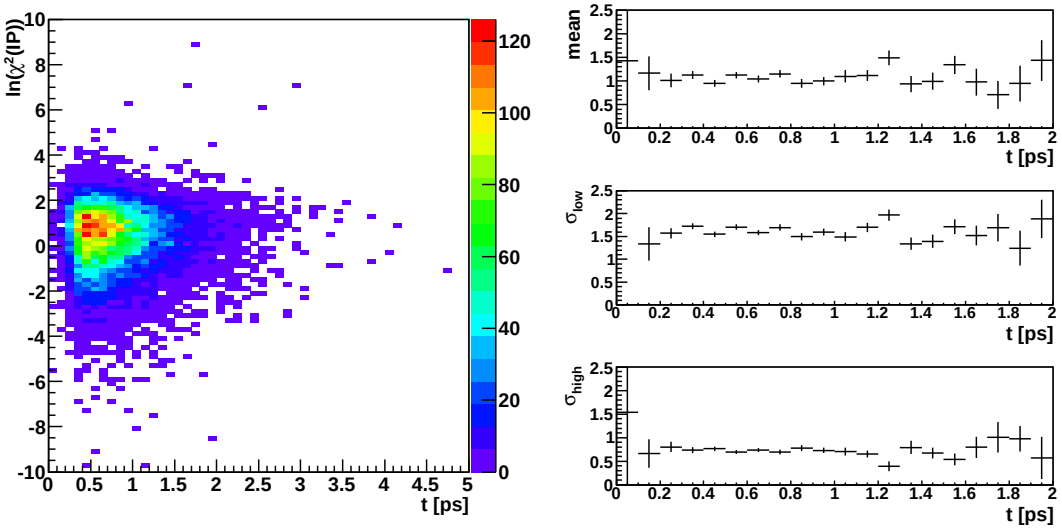


Figure 5.4: (left) The distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time for prompt D^0, from full Monte Carlo simulation. (right) The fitted values of the mean, and lower and upper σ of a bifurcated Gaussian in each bin of proper decay time.

σ_{low} and σ_{high} obtained from fitting a bifurcated Gaussian in each bin of proper decay time. As expected, no significant variation in these parameters is seen between bins. This justifies the use of a PDF that is independent of proper decay time. To fit the data sufficiently well a sum of two bifurcated Gaussians and one symmetric Gaussian, all with a common mean, is used. Thus the PDF for $\ln(\chi^2(IP_{\mathrm{D}^0}))$ for prompt D^0 is given by

$$\begin{aligned} f_{signal,IP}(x) =& f_{IP1} Bif.Gauss(x, \mu_{IP}, \sigma_{lowIP1}, \sigma_{highIP1}) \\ &+ f_{IP2} Bif.Gauss(x, \mu_{IP}, \sigma_{lowIP2}, \sigma_{highIP2}) \\ &+ (1 - f_{IP1} - f_{IP2}) Gauss(x, \mu_{IP}, \sigma_{IP3}). \end{aligned} \tag{5.10}$$

The distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time for secondary D^0, also obtained from simulation, is shown in figure 5.5. Here there is a clear dependence of the μ of the bifurcated Gaussian fit on proper decay time. A fit to the distribution of the μ values is also shown, using the parametrisation

$$\mu(t) = \mu_{IP0} + \mu_{IP1} \ln(1 + t/t_{IP0}). \tag{5.11}$$

This fits the data well, and so is used to parametrise the mean of the fit PDF as a function of proper decay time. There is some indication of a dependence of σ_{low} on proper decay time. However, to keep the PDF as simple as possible this is kept constant. The PDF of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ for secondary D^0 is also given by the sum of two bifurcated Gaussians and a symmetric Gaussian. The values of σ_{low} and σ_{high} in figure 5.5 are similar to those for prompt D^0. Consequently, the σ parameters are constrained to be the same as those for the PDF of

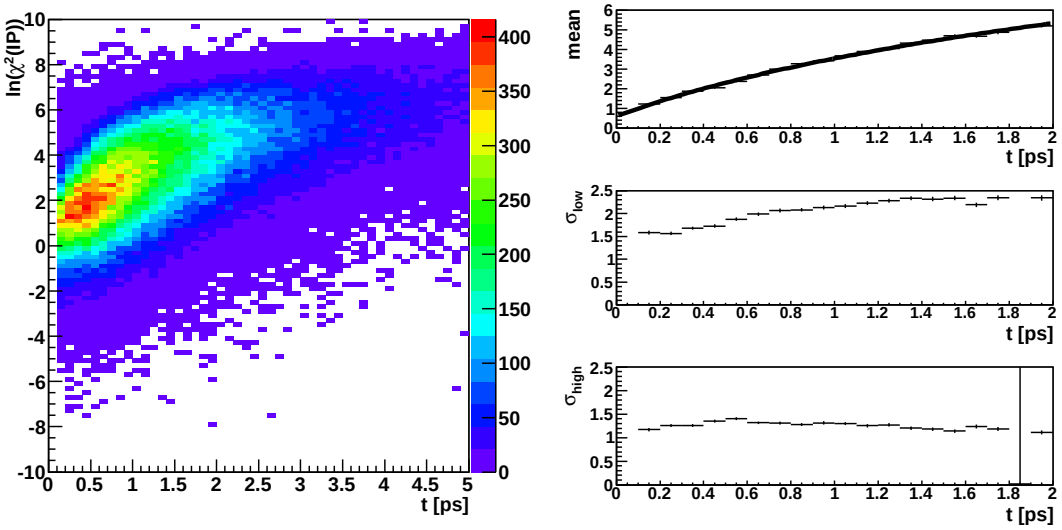


Figure 5.5: (left) The distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time for secondary D^0, from full Monte Carlo simulation. (right) The fitted values of the mean, and lower and upper σ of a bifurcated Gaussian in each bin of proper decay time.

prompt D^0, multiplied by a scaling factor. The relative fractions of the three components of the PDF are also constrained to be the same as in the PDF for prompt D^0. The PDF of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ for secondary D^0 is thus given by

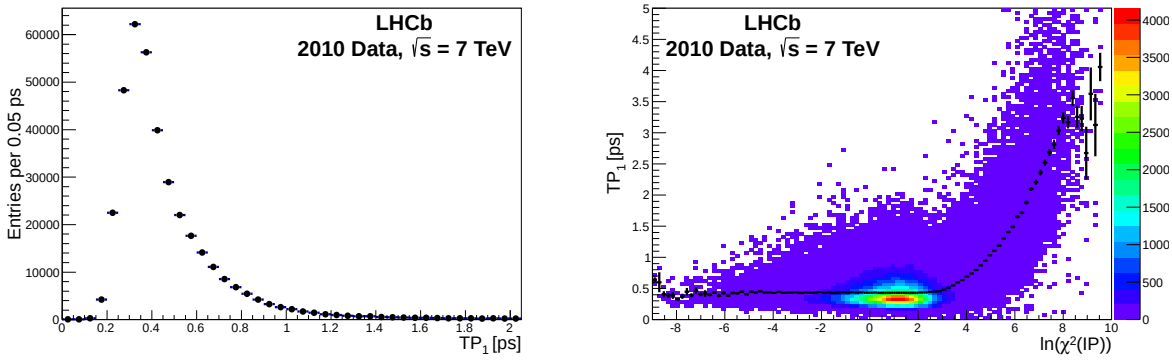
$$\begin{aligned} f_{secondary,IP}(x|t) =& f_{IP1} Bif.Gauss(x, \mu_{IP0} + \mu_{IP1}\ln(1+t/t_{IP0}), c_{IP\sigma}\sigma_{lowIP1}, c_{IP\sigma}\sigma_{highIP1}) \\ &+ f_{IP2} Bif.Gauss(x, \mu_{IP0} + \mu_{IP1}\ln(1+t/t_{IP0}), c_{IP\sigma}\sigma_{lowIP2}, c_{IP\sigma}\sigma_{highIP2}) \\ &+ (1 - f_{IP1} - f_{IP2}) Gauss(x, \mu_{IP0} + \mu_{IP1}\ln(1+t/t_{IP0}), c_{IP\sigma}\sigma_{IP3}). \end{aligned} \tag{5.12}$$

The value of t_{IP0} is fixed to be 180 fs again to aid the convergence of the fit. This value is obtained from the fit shown in figure 5.5.

5.3.1.3 The PDF of the Acceptance Variables

The final components required for the fit are the PDFs of the turning points, which has the general form $f_{class,TPs}(nTP, TP_1, ..., TP_n)$. Section 4.3.2 discussed the simplest scenario in which the distributions of turning points are the same for all classes of candidate, in which case their PDFs can be omitted from the fit PDF. To examine if this is the case for prompt and secondary D^0, plots are made of the distributions of the turning points as a function of $\ln(\chi^2(IP_{\mathrm{D}^0}))$. As shown in figure 5.3 the region of low $\ln(\chi^2(IP_{\mathrm{D}^0}))$ is dominated by prompt D^0, while secondary D^0 dominate at high $\ln(\chi^2(IP_{\mathrm{D}^0}))$.

Figure 5.6 shows the total distribution of TP_1 and its distribution as a function of $\ln(\chi^2(IP_{\mathrm{D}^0}))$, from the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ dataset. In the region where prompt D^0 dominate the average value



(a) **(b)**

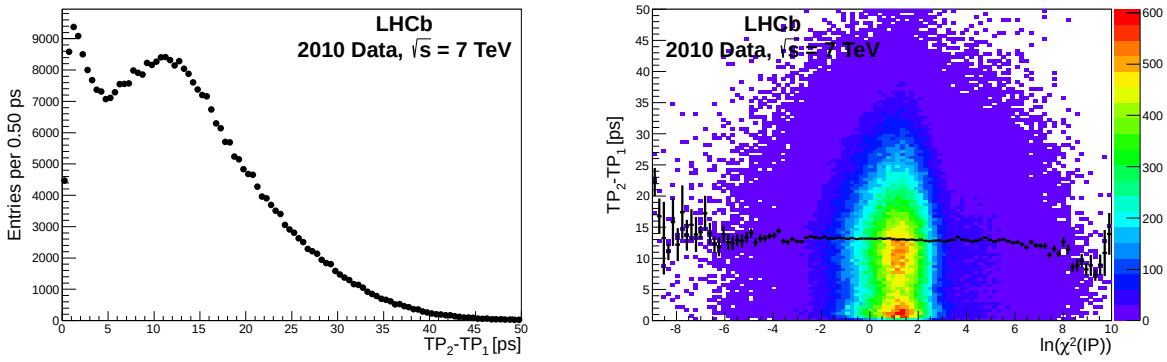
Figure 5.6: (a) The distribution of TP_1 and (b) the same distribution as a function of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ with the average value in each bin of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ overlaid, from 2010 $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ data.

of TP_1 is constant. At high $\ln(\chi^2(IP_{\mathrm{D}^0}))$, where secondary D^0 dominate, TP_1 tends to be much larger. As TP_1 and proper decay time are correlated (in that $t \geq TP_1$), and so are $\ln(\chi^2(IP_{\mathrm{D}^0}))$ and proper decay time for secondary D^0, this plot is insufficient to infer a dependence of TP_1 on $\ln(\chi^2(IP_{\mathrm{D}^0}))$. However, it does indicate that the distribution of TP_1 is very different for prompt and secondary D^0. Thus, separate PDFs of the turning points for prompt and secondary D^0 must be included in the fit PDF.

The PDF of the turning points, $f_{class,TPs}(nTP, TP_1, ..., TP_n)$, describes a large number of variables, and so is difficult to implement in the fit. In [94] this is circumvented by using a Fischer discriminant to transform the vector of turning point values into a single value, and a systematic error assigned for any bias this might introduce. An alternative approach, used here, is to examine the turning point distributions and simplify their PDFs accordingly.

Firstly, the value of TP_1 should be determined entirely by the candidate selection and the kinematics of the candidates, thus it is natural that TP_1 should follow a different distribution for prompt and secondary D^0. As discussed in section 4.3.2.1, the only reasons that additional turning points appear are due to having multiple PVs in the event, and the geometric acceptance of the VELO. These are independent of the candidate kinematics and selection criteria, making it a safe assumption that TP_1 is independent of nTP.

In addition, the following turning points can be assumed to be correlated only in that $TP_n \geq TP_{n-1}$. Thus, transforming the fit variables from the turning points to the difference between two consecutive turning points yields a set of uncorrelated fit variables, and the PDF



(a) **(b)**

Figure 5.7: (a) The distribution of $TP_2 - TP_1$ and (b) the same distribution as a function of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ with the average value in each bin of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ overlaid, from 2010 $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ data.

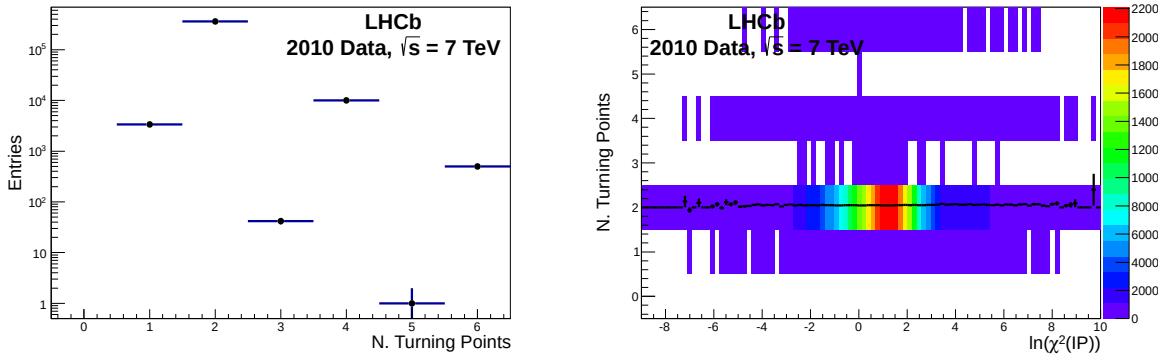
becomes

$$f_{class,TPs}(nTP, TP_1, ..., TP_n) = f_{class,nTP}(nTP) f_{class,TP1}(TP_1) \prod_{n=2}^{nTP} f_{class,\Delta TP_n}(TP_n - TP_{n-1}|nTP). \quad (5.13)$$

One cannot assume that the difference between turning points is independent of nTP. For example, in the case of only two turning points being found for a given candidate the difference between the first and second turning points is likely to be large; if four turning points are found this difference is likely to be smaller. Nonetheless, this is a much simpler PDF to implement in the fit.

One can further reason that as the distributions of the number of turning points and the differences between turning points are independent of the candidate kinematics they should be the same for prompt or secondary D^0. Figure 5.7 shows the total distribution of the difference between the first and second turning points and its distribution and mean value as a function of $\ln(\chi^2(IP_{\mathrm{D}^0}))$. No significant variation is seen, thus confirming the assumption that this distribution is the same for prompt and secondary D^0. Figure 5.8 also shows the distribution of the number of turning points. Again, this is constant as a function of $\ln(\chi^2(IP_{\mathrm{D}^0}))$, and so can be assumed to be the same for prompt and secondary D^0. Thus, only the PDF of the first turning point need be included in the fit PDF.

Unlike the proper decay time and mass, there are no physically motivated PDFs that can be used for these variables. The best approach is thus to extract the PDFs from the data themselves. A cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 1$ is applied to select a data sample of almost pure prompt D^0. Kernel density estimation, as described in section 4.1.3, is then used on this dataset to obtain a PDF of the first turning point for prompt D^0.



(a) **(b)**

Figure 5.8: (a) The distribution of the number of turning points and (b) the same distribution as a function of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ with the average value in each bin of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ overlaid, from 2010 $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ data.

Obtaining the PDF for secondary D^0 is rather more difficult. As described in section 5.3.1.2, parametric PDFs for the proper decay time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ are used for both prompt and secondary D^0. Thus, their shape is determined reasonably well before the fit by using the optimal parameter values obtained in a previous fit. As only the PDF of the first turning point need be considered for the fit, this PDF for secondary D^0 is the only unknown one. A rough first estimate of this PDF is obtained by using kernel density estimation on the first turning point values for the whole dataset.

Thus one has a PDF for each class for the set of variables $\underline{\mathbf{T}}$, $f_{class,T}(\underline{\mathbf{T}})$. Although the PDF of the first turning point for secondary D^0 is not correct at this time, these PDFs can still be used to calculate the probability of a candidate being prompt or secondary D^0

$$P(class|\underline{\mathbf{T}}_i) = \frac{f_{class}(\underline{\mathbf{T}}_i)P(class)}{\Sigma_{class} f_{class}(\underline{\mathbf{T}}_i)P(class)}, \tag{5.14}$$

where the PDF for a given class has the form of the right hand side of equation 5.5. Again, the term $f_{class,TPs}(nTP, TP_1, ..., TP_n)$ is simply replaced here by $f_{class,TP1}(TP_1)$, as its other components factor out of this calculation.

As is explained in [96], weighting each candidate by $P(class|\underline{\mathbf{T}}_i)$ for a specific class and histogramming one of the variables in $\underline{\mathbf{T}}$ yields the distribution of that variable for that class, assuming the PDFs used to calculate the weights are correct. If the PDFs are incorrect the distribution obtained will differ from the PDF used to describe the variable. Thus, calculating the weights, $P(secondary|\underline{\mathbf{T}}_i)$, and using them to weight the kernel functions when plotting the values of TP_1, a new PDF is obtained. If the initial PDF of TP_1 for secondary D^0 is correct, the new PDF is close to identical to the original. If not, the new PDF is different, but slightly closer to the correct distribution than the previous PDF. In this case, the new PDF

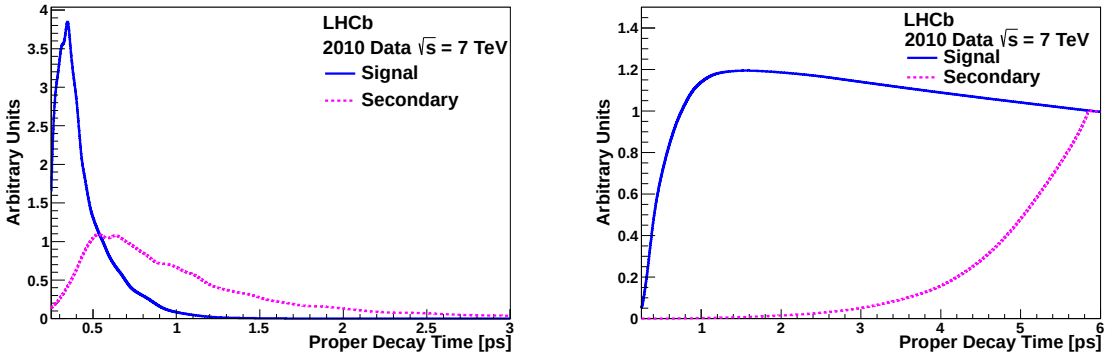


Figure 5.9: (a) The PDFs of the first turning point for prompt and secondary D^0, obtained from data using kernel density estimation. (b) The average acceptance as a function of proper decay time for prompt and secondary D^0, calculated using equation 5.15.

is used to recalculate the weights $P(secondary|\underline{\mathbf{T}}_i)$ and then used to plot TP_1 again. Thus, this process is iterated upon until the PDF used to calculate the weights and the one obtained using them agree sufficiently well. At this point one can be certain that the PDF describes the data well.

This assumes that the other fit PDFs used in calculating the weights describe the data well, prior to the fit. Thus, there is some dependence of the PDF of TP_1 obtained on the initial values of the parameters of the other fit PDFs. Provided these initial values are reasonably close to the final fitted values the PDF obtained will be sufficiently accurate, as it is still mostly defined by the distribution of the data. Any remaining influence of the initial fit parameter values is almost completely removed by a second iteration of the proper-decay-time fit, which is described in section 5.3.1.5. This iterative technique can be used for any PDF for any class, but only in the case that only one PDF for one variable is unknown.

Thus, separate PDFs of the first turning point for prompt and secondary D^0 are obtained. An example of the PDFs thus obtained, using $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ data, are shown in figure 5.9a. The PDFs of the first turning point for prompt and secondary $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$, extracted from data using this method, are shown in figure 5.9a. As expected, TP_1 takes much higher values for secondary D^0 than for prompt D^0.

As these PDFs are obtained using kernel density estimation they have no free parameters, and so remain constant throughout the fit procedure. The full list of parameters in the proper-decay-time fit is shown in table 5.7.

Parameter	Class	Description
f_{prompt}	Both	The fraction of prompt D^0 in the dataset
τ_{D^0}	Prompt	The effective lifetime of the signal
τ_1	Secondary	The first τ parameter of the proper-decay-time PDF - fixed to 270 fs
τ_2	Secondary	The second τ parameter of the proper-decay-time PDF
σ_t	Both	The σ of the Gaussian proper-decay-time resolution function - fixed to 50 fs
μ_{IP}	Prompt	The central value of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF - applies to both bifurcated Gaussians and the symmetric Gaussian
f_{IP1}	Both	The relative fraction of the first bifurcated Gaussian of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF
σ_{lowIP1}	Both	The lower σ of the first bifurcated Gaussian of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF
$\sigma_{highIP1}$	Both	The upper σ of the first bifurcated Gaussian of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF
f_{IP2}	Both	The relative fraction of the second bifurcated Gaussian of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF
σ_{lowIP2}	Both	The lower σ of the second bifurcated Gaussian of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF
$\sigma_{highIP2}$	Both	The upper σ of the second bifurcated Gaussian of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF
σ_{IP3}	Both	The σ of the symmetric Gaussian of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF
$c_{IP\sigma}$	Secondary	The scale factor between the σ parameters of prompt and secondary $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDFs
μ_{IP0}	Secondary	The central value of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF at $t = 0$ - applies to both bifurcated Gaussians and the symmetric Gaussian
μ_{IP1}	Secondary	Describes the rate of increase of the central value of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF as a function of t
t_{IP0}	Secondary	Describes the rate of increase of the central value of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF as a function of t - fixed to 180 fs

Table 5.7: The parameters of the proper-decay-time fit and their meaning. The class of candidate to which they apply is also shown; this can be prompt D^0, secondary D^0, or both.

5.3.1.4 Calculation of the Average Acceptance as a Function of Proper Decay Time

As different PDFs for the first turning point are required for prompt and secondary D^0 the simple method of calculating the average acceptance function by summing the acceptance functions for each individual candidate, discussed in section 4.3.2.3, cannot be used. An acceptance function obtained by doing so would contain contributions from the turning point distributions of the other classes of candidate, and so would not be accurate. Thus, the average acceptance function must be calculated for each class by integrating over the turning point PDFs specific to each class.

Using the form of the PDF of the turning points in equation 5.13 and that nTP takes integer values, the integral required to calculate the average acceptance function becomes

$$A_{class,avg}(t) = \Sigma_{nTP=2}^{nTP_{max}} \Bigg[P(nTP|class) \int_{-\infty}^{+\infty} \Bigg(\frac{\left(\Sigma_{n=1}^{nTP/2} \Theta(t - TP_{2n-1})\Theta(TP_{2n} - t)\right)}{\Sigma_{n=1}^{nTP/2} \int_{TP_{2n-1}}^{TP_{2n}} f_{class,t}(t')\mathrm{d}t'} f_{class,TP1}(TP_1) \prod_{n=2}^{nTP} f_{class,\Delta TP_n}(TP_n - TP_{n-1}|nTP) \Bigg) \mathrm{d}TP_1 ... \mathrm{d}TP_{nTP} \Bigg]. \tag{5.15}$$

Examples of the PDF of the first turning point, $f_{class,TP1}(TP_1)$, are shown for prompt and secondary $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ in figure 5.9a. Note that, although the PDFs of the differences between turning points are not implemented in the fit, they need to be known to properly calculate the average acceptance function. Thus, normalised histograms, made using the full dataset, are used to estimate their distributions.

As the PDFs of the turning points and their separation aren't analytical functions this integral can only be evaluated numerically. Taking N bins in TP_1 of width δTP_1, and N bins in $(TP_2 - TP_1)$ of width δTP_{2-1} one can define

$$TP_{1,i} = TP_{1,0} + i\delta TP_1, \quad \text{and} \tag{5.16}$$

$$TP_{2,ij} = TP_{1,0} + i\delta TP_1 + j\delta TP_{2-1}. \tag{5.17}$$

Then one can express the term for $nTP = 2$ as a sum

$$P(nTP = 2|class)\Sigma_{i=0}^{N}\Sigma_{j=0}^{N} \Bigg(\frac{(\Theta(t - TP_{1,i})\Theta(TP_{2,ij} - t))}{\int_{TP_{1,i}}^{TP_{2,ij}} f_{class,t}(t')\mathrm{d}t'} f_{class,TP1}(TP_{1,i})\delta TP_1 f_{class,\Delta TP_2}(TP_{2,ij} - TP_{1,i}|nTP = 2)\delta TP_{2-1} \Bigg). \tag{5.18}$$

This is essentially the total sum of the per-candidate acceptance functions for each value of TP_1 and TP_2 weighted by the probability of finding that number of turning points and those turning point values. Analogous calculations are performed for higher values of nTP, up to its maximum value of six.

Each term for each value of nTP thus requires calculating and summing N^{nTP} acceptance functions. For $nTP = 2$ using $N = 1000$ requires 10^6 separate calculations, which is manageable. However, using $N = 1000$ for $nTP = 4$ one would have 10^{12} calculations, and 10^{18} for $nTP = 6$, which would require a prohibitive amount of CPU time. Thus N is scaled according to the value of nTP in order to be able to perform such a calculation within reasonable time limits. This results in requiring very small N for large nTP. However, as shown in figure 5.8, $P(nTP|class)$ is very small for high nTP. Thus, the calculation of the average acceptance is dominated by the $nTP = 2$ term, and higher nTP terms contribute very little. Hence, using small N for high nTP does not significantly reduce the accuracy of the average acceptance function thus obtained.

An example of the acceptance functions calculated in this manner for prompt and secondary $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ are shown in figure 5.9b. For prompt D^0 the acceptance rate is very small at low proper decay times, as expected from the use of minimum χ^2_{IP} cuts on the D^0 daughter tracks. The presence of multiple turning points, due to multiple PVs in the events, causes the acceptance to peak at moderate proper decay times and drop slightly at high proper decay times. This effect is small as the average number of PVs reconstructed by LHCb during the relevant data-taking period is only $\sim$2.1. This is consistent with expectations, as discussed in section 4.3.2.1. Secondary D^0 have very low acceptance at low and moderate proper decay times, as the first turning point tends to take much larger values.

5.3.1.5 Suppression of Secondary D^0

The difficulties in obtaining a sufficiently accurate description of secondary D^0 result in a large systematic uncertainty on the effective lifetimes determined using the full dataset. Consequently, an additional fit is performed on a subset of the data after the fit to the full dataset. The fraction of secondary D^0 is suppressed for this second fit by applying a cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$. The parameters of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDFs are fixed in the second fit to the values obtained in the first fit using the full dataset. The fraction of prompt D^0 in the reduced dataset is calculated by evaluating the integral of the fit PDFs for prompt and secondary D^0 up to the cut value. This fraction is then also fixed. Thus, the only free parameters in the second fit are τ_{D^0} and τ_2.

Due to the correlation between $\ln(\chi^2(IP_{\mathrm{D}^0}))$ and proper decay time for secondary D^0, the cut $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$ introduces an average acceptance as a function of proper decay

	D^0	$\overline{\mathrm{D}}^0$	Untagged
$\mathrm{K}^+\mathrm{K}^-$	15,234	15,247	30,481
$\mathrm{K}\pi$	111,212	114,898	226,110

Table 5.8: The number of candidates used in the fits to extract y_{CP} and A_Γ, including backgrounds, after requiring $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$.

time for secondary D^0. This is given by the integral of the PDF of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ up to the cut value:

$$A_{sec}(t|\chi^2_{max}(IP_{\mathrm{D}^0})) = \int_{-\infty}^{\ln(\chi^2_{max}(IP_{\mathrm{D}^0}))} f_{sec,IP}(x|t)\mathrm{d}x. \tag{5.19}$$

Thus the proper decay time PDF for secondary D^0 becomes

$$f_{sec,t}(t|\chi^2_{max}(IP_{\mathrm{D}^0})) = \frac{1}{N} A_{sec}(t|\chi^2_{max}(IP_{\mathrm{D}^0}))f_{sec,t}(t), \tag{5.20}$$

where the normalisation is given by

$$N = \int_{-\infty}^{+\infty} A_{sec}(t|\chi^2_{max}(IP_{\mathrm{D}^0}))f_{sec,t}(t)\mathrm{d}t. \tag{5.21}$$

Such a cut also modifies the distribution of the first turning point for secondary D^0. This PDF is thus re-evaluated on the reduced dataset using the method discussed in section 5.3.1.3, with only one iteration. It is produced using kernel density estimation, and weighting each candidates contribution by $P(secondary|\underline{\mathbf{T}}_i)$, as determined by the first fit and defined in equation 5.14. This has the consequence that any dependence of the PDF of TP_1 for secondary D^0 on the initial values of the fit parameters, prior to the first fit, is almost completely eliminated. The values of τ_{D^0} determined by this second fit for each decay channel are then used to calculate y_{CP} and A_Γ (after the correction for mis-tagged D^0 in the case of A_Γ).

The number of candidates passing the trigger and offline selection criteria, detailed in section 5.1, are shown in table 5.3. These are used in the first iteration of the fit. The number of candidates also fulfilling the requirement of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$, that are used in the second fit iteration, are shown in table 5.8.

5.3.1.6 Blinding

During the development of the final fit methodology the lifetimes of the $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ datasets were blinded. This was done to ensure that no user imposed bias, either conscious or unconscious, could be introduced to the observed values of A_Γ and y_{CP} before the final fit methodology was defined. The blinding was achieved by scaling the lifetimes output by the fit by an unknown factor between 0.97 and 1.03. Only once the full fit procedure was agreed upon were the blinding factors removed.

5.3.1.7 Future Considerations

Due to the complicated nature of the full fit PDF it is possible that the approximations made here may prove not to be exact when examining larger datasets in the future.

One minor issue exists due to the geometric acceptance of the VELO, or in the presence of a maximum cut on the proper decay time. In this case the PDF of the difference between the last and second last turning points is modified as the last turning point always falls in roughly the same place (or exactly the same place when applying a maximum cut). This would affect the PDFs of the differences between turning points for secondary D^0 much more, as the value of TP_1 takes much larger values. However, as the contribution of secondary D^0 to the dataset is suppressed so strongly in the second iteration of the fit, this effect is sufficiently small as to be neglected for the measurements presented here. A relevant systematic uncertainty is nonetheless calculated in section 6.3.5.

Another consideration that must be made comes from properly examining the PDF of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ in equation 4.17. In the previous section it was assumed that $\ln(\chi^2(IP_{\mathrm{D}^0}))$ depends only on t for secondary D^0; however, this PDF is in fact conditional on the full set of proper-decay-time fit variables, $\{t, nTP, TP_1, ..., TP_n\}$. As, during the swimming algorithm, PVs are moved along the direction of the momentum vector of the D^0 the IP, and $\chi^2(IP_{\mathrm{D}^0})$, of the D^0 does not change with the distance swum. However, this does not rule out the possibility of correlation between the $\chi^2(IP_{\mathrm{D}^0})$ and the values of the turning points. Intuitively, one might expect that a D^0 candidate with a very large $\chi^2(IP_{\mathrm{D}^0})$, such as those for secondary D^0 at high proper decay times, may well be inclined to have a larger value for TP_1. A larger $\chi^2(IP_{\mathrm{D}^0})$ implies a larger pointing angle; at lower proper decay times the pointing angle will get larger still, and so may well cause the candidate not to pass the selection, resulting in a large value for TP_1. Hence a correlation between $\chi^2(IP_{\mathrm{D}^0})$ and TP_1 is plausible, in addition to the dependence of $\chi^2(IP_{\mathrm{D}^0})$ on proper decay time for secondary D^0.

As shown by figure 5.6 the value of TP_1 appears uncorrelated to $\ln(\chi^2(IP_{\mathrm{D}^0}))$ for prompt D^0, and so such considerations only affect secondary D^0 (and potentially combinatorial backgrounds in future). Thus, as secondary D^0 contribute so little to the final fit dataset, it is reasonable to neglect such effects for the measurements presented here. Again, a corresponding systematic uncertainty is applied in section 6.3.6. However, such effects will need to be studied carefully when larger datasets are examined in future.

5.3.2 Measurements of Effective Lifetimes

The simultaneous fit to the proper-decay-time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distributions, detailed in section 5.3.1, is applied to each of the datasets to determine the effective lifetime of the signal. The first iteration of the fit, which is used to determine the parameters of the PDFs of $\ln(\chi^2(IP_{\mathrm{D}^0}))$, is performed using the statistics shown in table 5.3. The signal lifetime (and τ_2 parameter for secondary D^0) is determined in the second iteration of the fit, after the application of the cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$, the statistics for which are shown in table 5.8. Table 5.9 shows the fitted values and their statistical errors for all parameters in the fits to the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$ datasets. Table 5.10 shows the same results for $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-$. The results of the fits to the combined, untagged datasets are shown in table 5.11. No significant correlation between τ_{D^0} and any of the other fit parameters is observed in either iteration of the proper-decay-time fit on any dataset.

The projections of the corresponding fits onto the proper-decay-time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ axes are shown for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$ in figure 5.10. The pull of the fit in each bin is shown below each plot. For secondary D^0 the projection onto the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ axis is achieved by integrating the PDF of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ multiplied by the average proper-decay-time PDF over the range of proper decay time in the fit. The average proper-decay-time PDF for each class is obtained by integrating over the range of turning point values and the number of turning points, up to the maximum of six, as described by equation 5.15. Examples of the acceptance functions thus obtained are shown in figure 5.9b.

Figure 5.11 shows the distributions of the data and the pull values of the fits in the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ and proper-decay-time plane, for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$ after the final iteration of the fit. The pulls are evenly distributed about zero, showing that the fitted PDFs describe the data well. Figures 5.12 and 5.13 show the fit projections and their pulls for $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$, $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-$ and untagged $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ and $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$. The pulls of the projections are also generally distributed evenly about zero. With the higher statistics of $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ some small regions where the pulls are consistently positive or negative are apparent, for example the region $2 < \ln(\chi^2(IP_{\mathrm{D}^0})) < 6$ in figure 5.13c. This implies some small inaccuracies in the fit PDFs, though there are no apparent structures to the pulls in the region of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$, in which the final fit is performed to extract the effective lifetimes. Thus, it is clear that the average acceptance as a function of proper decay time is reproduced accurately using the data obtained by the swimming algorithm, and the parametrisation of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distribution is sufficiently good. Any systematic uncertainties resulting from inaccuracies in the fit PDFs are evaluated in section 6.3.

A summary of the effective lifetimes determined on each dataset is shown in table 5.12. The lifetimes for the tagged $\mathrm{K}^+\mathrm{K}^-$ datasets after applying the correction for the bias from

$D^0 \to K^-\pi^+$ Proper-Decay-Time Fit Results	
Parameter	Fitted Value
f_{prompt}	0.93019 ± 0.00086
τ_1	$0.270\,\text{ps}$
μ_{IP}	1.306 ± 0.014
f_{IP1}	0.557 ± 0.016
σ_{lowIP1}	1.652 ± 0.041
$\sigma_{highIP1}$	0.724 ± 0.022
f_{IP2}	0.120 ± 0.020
σ_{lowIP2}	2.777 ± 0.080
$\sigma_{highIP2}$	1.362 ± 0.039
σ_{IP3}	0.922 ± 0.018
$c_{IP\sigma}$	0.929 ± 0.011
μ_{IP0}	0.558 ± 0.071
μ_{IP1}	1.969 ± 0.028
t_{IP0}	$0.180\,\text{ps}$
f_{prompt}	0.99518
τ_2	$0.59 \pm 0.16\,\text{ps}$
τ_{D^0}	$410.6 \pm 1.3\,\text{fs}$

(a)

$\bar{D}^0 \to K^+\pi^-$ Proper-Decay-Time Fit Results	
Parameter	Fitted Value
f_{prompt}	0.93055 ± 0.00085
τ_1	$0.270\,\text{ps}$
μ_{IP}	1.298 ± 0.013
f_{IP1}	0.571 ± 0.016
σ_{lowIP1}	1.684 ± 0.040
$\sigma_{highIP1}$	0.747 ± 0.022
f_{IP2}	0.102 ± 0.018
σ_{lowIP2}	2.888 ± 0.088
$\sigma_{highIP2}$	1.401 ± 0.042
σ_{IP3}	0.911 ± 0.018
$c_{IP\sigma}$	0.934 ± 0.011
μ_{IP0}	0.344 ± 0.070
μ_{IP1}	2.066 ± 0.028
t_{IP0}	$0.180\,\text{ps}$
f_{prompt}	0.99498
τ_2	$0.84 \pm 0.19\,\text{ps}$
τ_{D^0}	$409.9 \pm 1.3\,\text{fs}$

(b)

Table 5.9: The fitted values and their statistical errors of the parameters of the simultaneous fit to the proper decay time and $\ln(\chi^2(IP_{D^0}))$ distributions, for (a) $D^0 \to K^-\pi^+$ and (b) $\bar{D}^0 \to K^+\pi^-$. The values of the parameters in the last three rows are those determined in the second fit, after applying the cut of $\ln(\chi^2(IP_{D^0})) < 2$. Parameters without errors are fixed in the fit. The definitions of the parameters are given in table 5.7 on p. 139. No significant correlation between τ_{D^0} and any of the other fit parameters is observed in either iteration of the proper-decay-time fit. The corresponding fit projections are shown in figure 5.10.

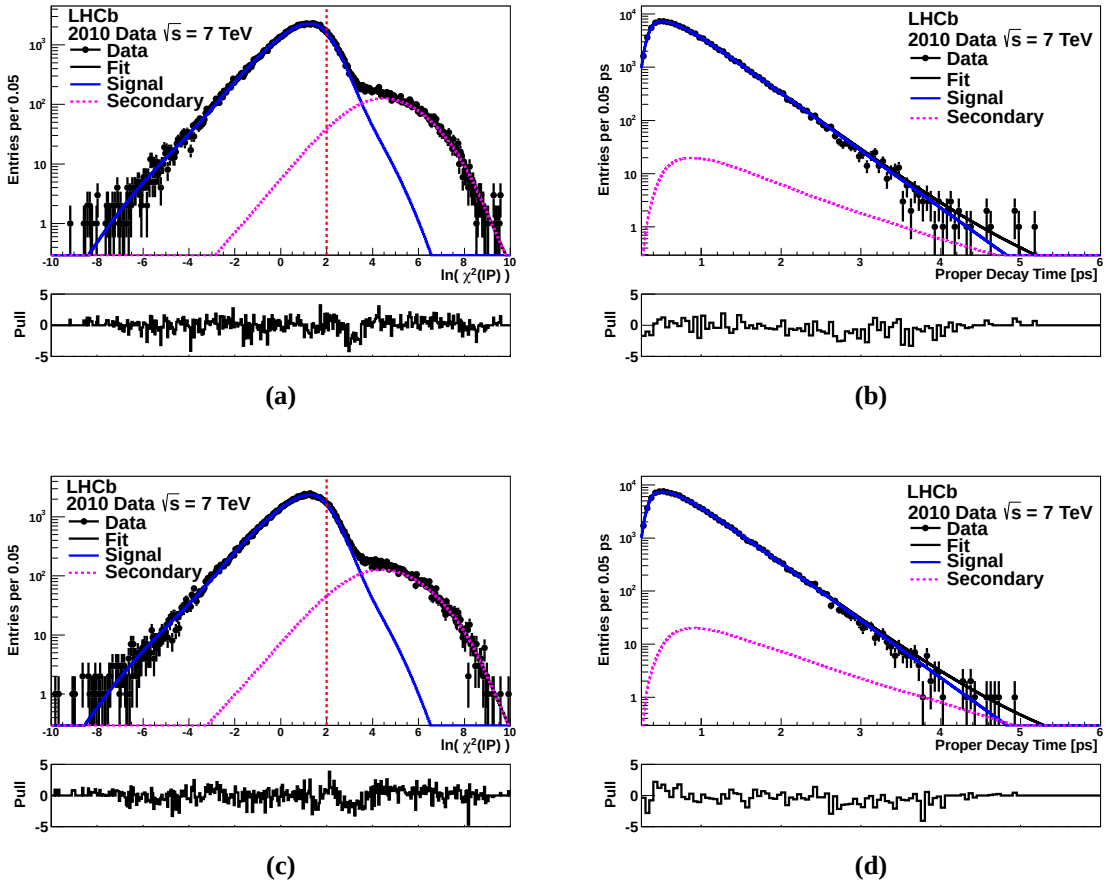


Figure 5.10: (a) The projection onto the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ axis of the first iteration of the simultaneous fit to the proper decay time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distributions for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$. The red dashed line shows the cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$ used for the second iteration of the fit. (b) The projection onto the proper decay time axis of the results of the second iteration of the fit, using candidates left of the red dashed line in (a). (c) and (d) show the same for $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$. The pull of the fit in each bin is shown below each plot. The corresponding fit results are shown in table 5.9.

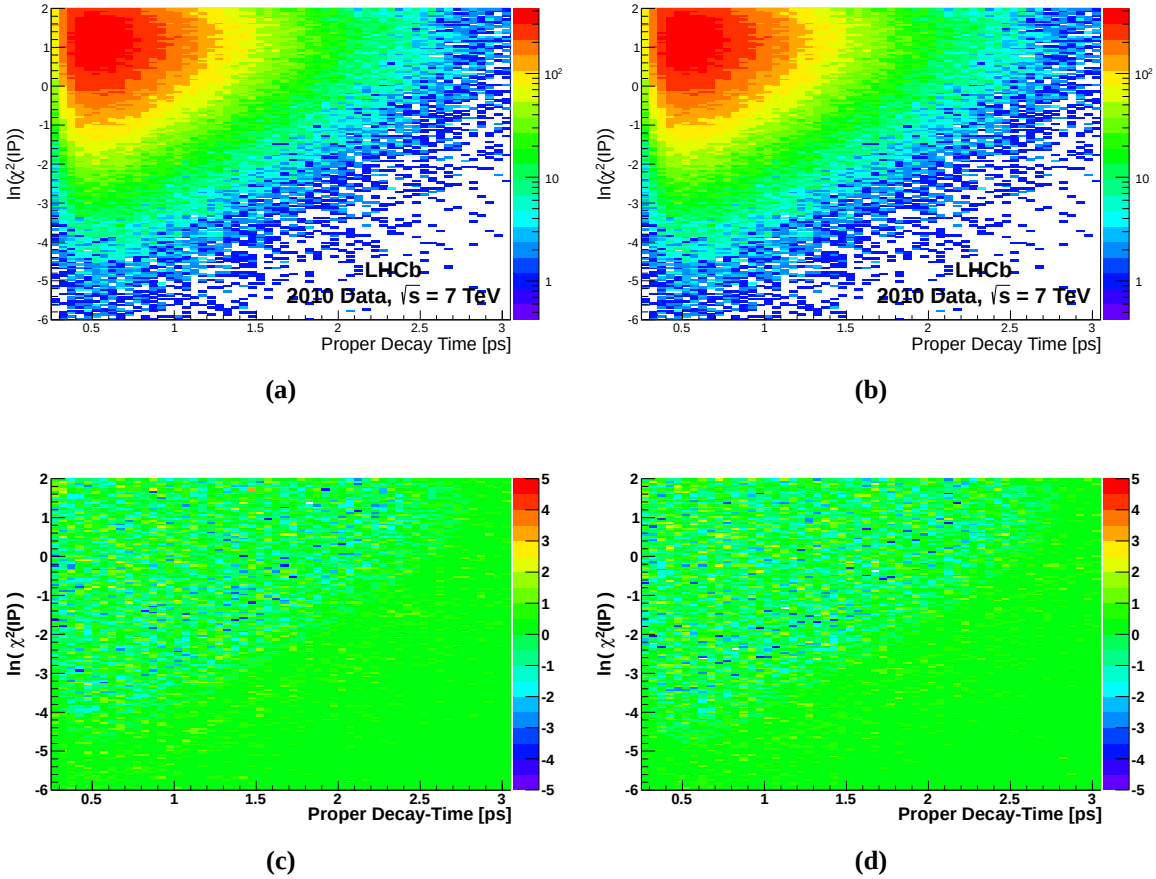


Figure 5.11: Distribution of $\ln(\chi^2(IP_{\mathrm{D}^0}))$ as a function of proper decay time from the subset of the data used in the final iteration of the proper-decay-time fit, for (a) $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and (b) $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$. (c) and (d) show the pull values (data minus fit, divided by the statistical error) in each bin.

mis-tagged D^0 are also shown. These are used in the calculation of A_Γ. The corrections are performed using equations 5.7a and 5.7b with the mis-tag rates shown in equations 5.8a and 5.8b. The systematic uncertainties on the lifetimes themselves have not been estimated and would require careful consideration for these measurements to significantly improve the knowledge of the absolute D^0 lifetime. Hence, the lifetimes are quoted with statistical uncertainties only. These are used to calculate the values of and statistical uncertainties on y_{CP} and A_Γ. Many of the sources of systematic uncertainty on the absolute lifetimes are expected to cancel in the calculation of y_{CP} and A_Γ.

As a cross-check a 'pseudo' A_Γ measurement is also made using $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$, defined analogously to A_Γ as

$$A_\Gamma^{\mathrm{K}\pi,\mathit{eff}} = \frac{\tau_{\mathrm{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-) - \tau_{\mathrm{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)}{\tau_{\mathrm{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-) + \tau_{\mathrm{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)}. \tag{5.22}$$

Under CPT invariance the D^0 and $\overline{\mathrm{D}}^0$ should have identical effective lifetimes, thus $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ should yield a result consistent with zero. This also means that mis-tagged D^0 do not bias the measured effective lifetimes, and so no mis-tag correction is applied when calculating $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$. This provides a strong cross-check on the fit method as it exploits the significantly higher statistics in the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ channel compared to $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$.

The effective lifetimes shown in table 5.12 give the following results, showing only statistical uncertainties:

$$A_\Gamma^{\mathrm{K}\pi,\mathit{eff}} = (-0.9 \pm 2.2) \times 10^{-3}, \tag{5.23a}$$

$$A_\Gamma = (-5.9 \pm 5.9) \times 10^{-3}, \tag{5.23b}$$

$$y_{CP} = (5.5 \pm 6.3) \times 10^{-3}. \tag{5.23c}$$

As mentioned in section 5.3.1.6 the measurements of y_{CP} and A_Γ were blinded while the fit methodology was being developed. Further cross-checks on these measurements are shown in section 6.1, and the determination of their systematic uncertainties discussed in section 6.3.

5.4 Conclusions

This chapter presented the measurements of y_{CP} and A_Γ, the motivation and methodology for which has been discussed in chapters 1 and 4. Section 5.1 presented the datasets used and the selection criteria applied to the reconstructed D^0 candidates. The data comprises $28.0 \pm 2.8\,\mathrm{pb}^{-1}$ collected during the 2010 run, and yields 39,263 untagged $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ candidates and 286,159 untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ candidates after the selection is applied.

$D^0 \to K^+K^-$ Proper-Decay-Time Fit Results	
Parameter	Fitted Value
f_{prompt}	0.9231 ± 0.0024
τ_1	$0.270\,\text{ps}$
μ_{IP}	1.419 ± 0.044
f_{IP1}	0.579 ± 0.043
σ_{lowIP1}	1.64 ± 0.13
$\sigma_{highIP1}$	0.652 ± 0.044
f_{IP2}	0.152 ± 0.087
σ_{lowIP2}	2.62 ± 0.23
$\sigma_{highIP2}$	1.23 ± 0.10
σ_{IP3}	0.977 ± 0.047
$c_{IP\sigma}$	0.922 ± 0.028
μ_{IP0}	0.91 ± 0.18
μ_{IP1}	1.854 ± 0.072
t_{IP0}	$0.180\,\text{ps}$
f_{prompt}	0.9952
τ_2	$0.47 \pm 0.22\,\text{ps}$
τ_{D^0}	$410.4 \pm 3.4\,\text{fs}$

(a)

$\bar{D}^0 \to K^+K^-$ Proper-Decay-Time Fit Results	
Parameter	Fitted Value
f_{prompt}	0.9256 ± 0.0023
τ_1	$0.270\,\text{ps}$
μ_{IP}	1.260 ± 0.035
f_{IP1}	0.539 ± 0.045
σ_{lowIP1}	1.798 ± 0.066
$\sigma_{highIP1}$	0.803 ± 0.076
f_{IP2}	0.062 ± 0.013
σ_{lowIP2}	3.000 ± 0.079
$\sigma_{highIP2}$	1.714 ± 0.094
σ_{IP3}	0.956 ± 0.038
$c_{IP\sigma}$	0.924 ± 0.028
μ_{IP0}	0.29 ± 0.18
μ_{IP1}	2.057 ± 0.073
t_{IP0}	$0.180\,\text{ps}$
f_{prompt}	0.9947
τ_2	$0.27 \pm 0.30\,\text{ps}$
τ_{D^0}	$405.7 \pm 3.4\,\text{fs}$

(b)

Table 5.10: The fitted values and their statistical errors of the parameters of the simultaneous fit to the proper decay time and $\ln(\chi^2(IP_{D^0}))$ distributions, for (a) $D^0 \to K^+K^-$ and (b) $\bar{D}^0 \to K^+K^-$. The values of the parameters in the last three rows are those determined in the second fit, after applying the cut of $\ln(\chi^2(IP_{D^0})) < 2$. Parameters without errors are fixed in the fit. The definitions of the parameters are given in table 5.7 on p. 139. No significant correlation between τ_{D^0} and any of the other fit parameters is observed in either iteration of the proper-decay-time fit. The corresponding fit projections are shown in figure 5.12.

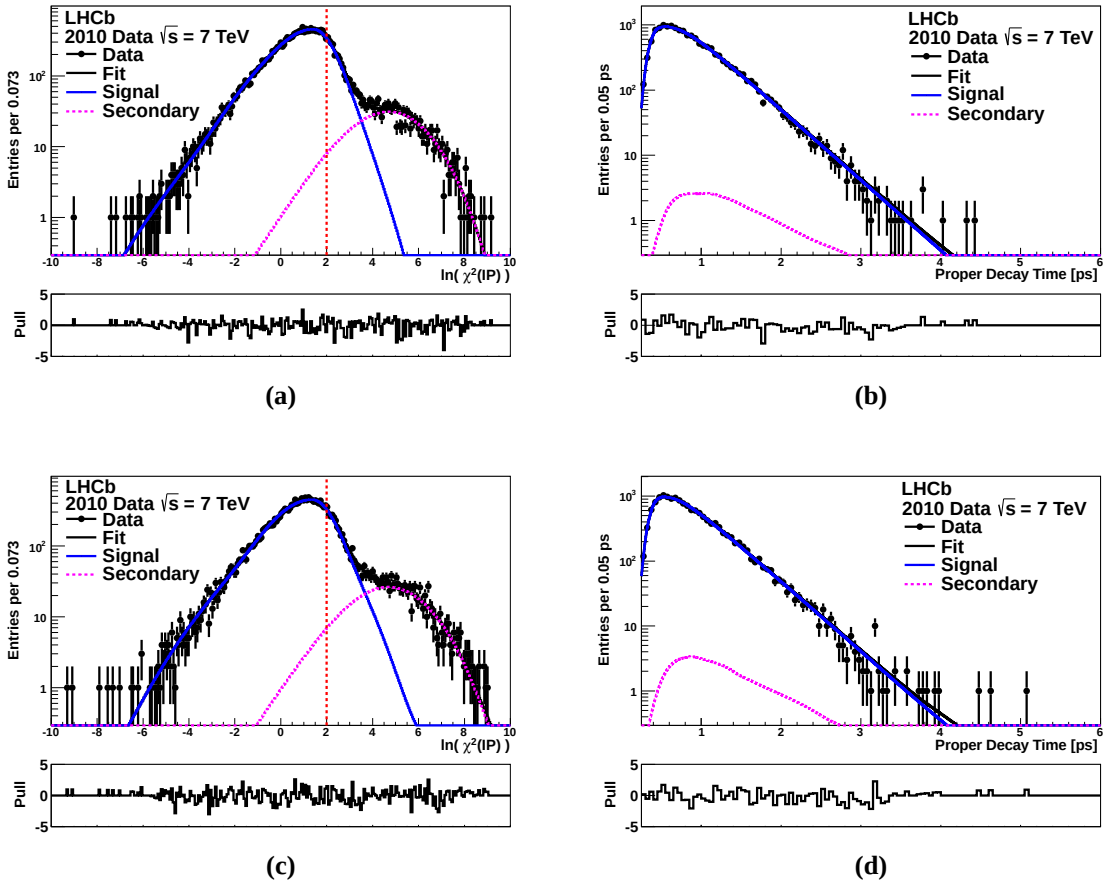


Figure 5.12: (a) The projection onto the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ axis of the first iteration of the simultaneous fit to the proper decay time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distributions for $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$. The red dashed line shows the cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$ used for the second iteration of the fit. (b) The projection onto the proper decay time axis of the results of the second iteration of the fit, using candidates left of the red dashed line in (a). (c) and (d) show the same for $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-$. The pull of the fit in each bin is shown below each plot. The corresponding fit results are shown in table 5.10.

Untagged $D^0 \to K^+K^-$ Proper-Decay-Time Fit Results	
Parameter	Fitted Value
f_{prompt}	0.9245 ± 0.0017
τ_1	$0.270\,\text{ps}$
μ_{IP}	1.328 ± 0.026
f_{IP1}	0.557 ± 0.026
σ_{lowIP1}	1.789 ± 0.038
$\sigma_{highIP1}$	0.723 ± 0.036
f_{IP2}	0.0639 ± 0.0097
σ_{lowIP2}	3.00 ± 0.37
$\sigma_{highIP2}$	1.567 ± 0.067
σ_{IP3}	0.976 ± 0.016
$c_{IP\sigma}$	0.922 ± 0.020
μ_{IP0}	0.60 ± 0.13
μ_{IP1}	1.949 ± 0.053
t_{IP0}	$0.180\,\text{ps}$
f_{prompt}	0.9951
τ_2	$0.39 \pm 0.18\,\text{ps}$
τ_{D^0}	$408.0 \pm 2.4\,\text{fs}$

(a)

Untagged $D^0 \to K^-\pi^+$ Proper-Decay-Time Fit Results	
Parameter	Fitted Value
f_{prompt}	0.93038 ± 0.00060
τ_1	$0.270\,\text{ps}$
μ_{IP}	1.3014 ± 0.0096
f_{IP1}	0.564 ± 0.011
σ_{lowIP1}	1.669 ± 0.029
$\sigma_{highIP1}$	0.736 ± 0.015
f_{IP2}	0.110 ± 0.013
σ_{lowIP2}	2.832 ± 0.059
$\sigma_{highIP2}$	1.382 ± 0.028
σ_{IP3}	0.917 ± 0.012
$c_{IP\sigma}$	0.9322 ± 0.0080
μ_{IP0}	0.450 ± 0.049
μ_{IP1}	2.018 ± 0.020
t_{IP0}	$0.180\,\text{ps}$
f_{prompt}	0.99507
τ_2	$0.72 \pm 0.13\,\text{ps}$
τ_{D^0}	$410.24 \pm 0.90\,\text{fs}$

(b)

Table 5.11: The fitted values and their statistical errors of the parameters of the simultaneous fit to the proper decay time and $\ln(\chi^2(IP_{D^0}))$ distributions, for (a) untagged $D^0 \to K^+K^-$ and (b) untagged $D^0 \to K^-\pi^+$. The values of the parameters in the last three rows are those determined in the second fit, after applying the cut of $\ln(\chi^2(IP_{D^0})) < 2$. Parameters without errors are fixed in the fit. The definitions of the parameters are given in table 5.7 on p. 139. No significant correlation between τ_{D^0} and any of the other fit parameters is observed in either iteration of the proper-decay-time fit. The corresponding fit projections are shown in figure 5.13.

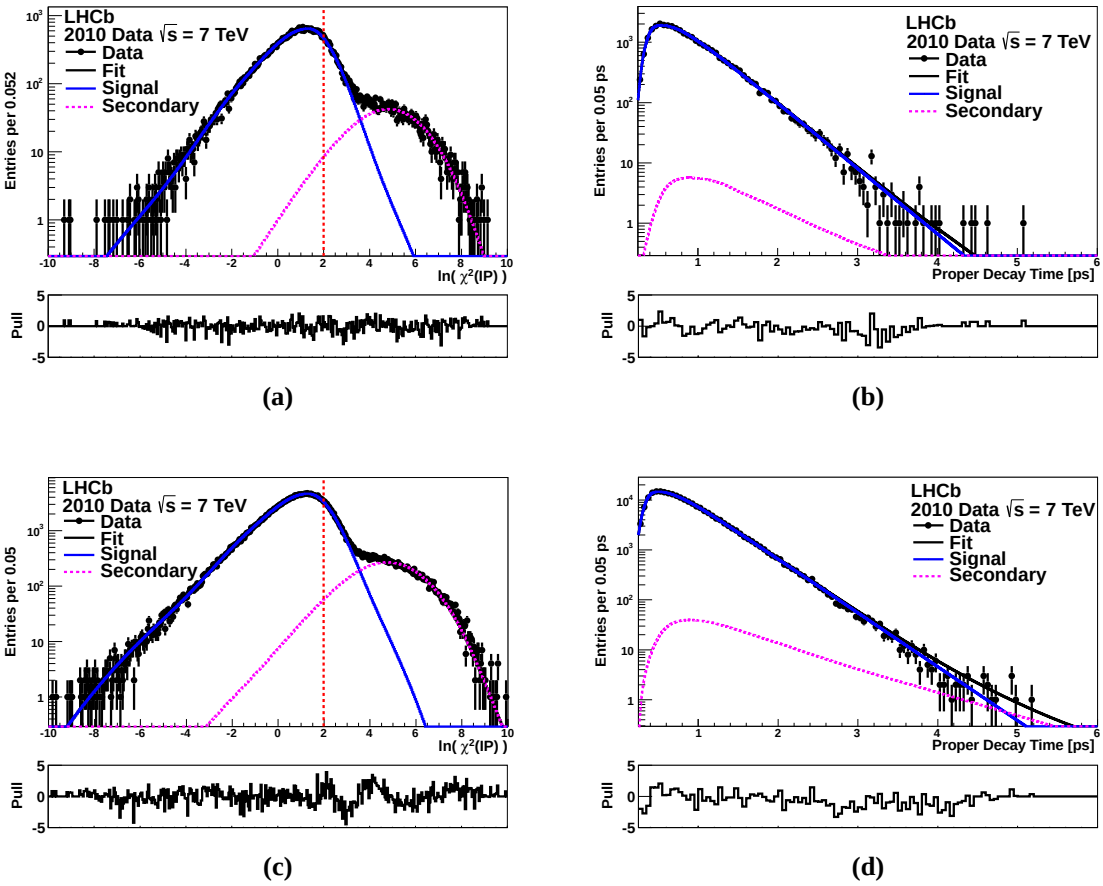


Figure 5.13: (a) The projection onto the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ axis of the first iteration of the simultaneous fit to the proper decay time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distributions for untagged $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$. The red dashed line shows the cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$ used for the second iteration of the fit. (b) The projection onto the proper decay time axis of the results of the second iteration of the fit, using candidates left of the red dashed line in (a). (c) and (d) show the same for untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$. The pull of the fit in each bin is shown below each plot. The corresponding fit results are shown in table 5.11.

Decay	D^0	$\overline{\mathrm{D}}^0$	Untagged
$\mathrm{K}^+\mathrm{K}^-$	410.4 ± 3.4 fs	405.7 ± 3.4 fs	408.0 ± 2.4 fs
$\mathrm{K}^+\mathrm{K}^-$ Mis-tag Corrected	410.5 ± 3.4 fs	405.6 ± 3.4 fs	
$\mathrm{K}\pi$	410.6 ± 1.3 fs	409.9 ± 1.3 fs	410.24 ± 0.90 fs

Table 5.12: A summary of the effective lifetimes determined for each decay channel and their statistical errors. The lifetimes for the tagged $\mathrm{K}^+\mathrm{K}^-$ final states are shown both before and after applying the correction for the bias from mis-tagged candidates.

The specifics of the fit methodology, discussed for more general cases in chapter 4, were described for the mass fits in section 5.2.1. The results of the mass fits to determine the fractions of combinatorial backgrounds and randomly-tagged D^0 were then shown in section 5.2.2. Due to the very small number of candidates accepted in the side-band regions, mass fits to the $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ datasets cannot converge stably, and so only results from the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ datasets were presented.

The specifics of the fits to extract the effective lifetimes were then detailed in section 5.3.1. This includes the exact parametrisation of the PDFs used in the fit, and the methods for extracting the distributions of the acceptance variables, determined by the swimming algorithm, from the data. Due to very limited statistics in the mass side-bands, combinatorial backgrounds are neglected in the fit to extract the effective lifetimes. The effects of mis-tagged D^0 are corrected after the lifetime fits have been performed using the random-tag rates determined by the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ mass fits. An additional fit iteration is also added after the application of a cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$, in order to suppress the component of secondary D^0 in the data. This reduces the final statistics to 30,481 untagged $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ candidates and 226,110 untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ candidates.

The results of the simultaneous fits to the proper-decay-time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distributions for each of $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$, $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-$, $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$, $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-$, and untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ were then shown in section 5.3.2. The fit PDFs and the average acceptance rate as a function of proper decay time determined by the swimming algorithm were shown to be accurate. A measurement of $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, defined analogously to A_Γ but using $\tau_{\mathrm{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ and $\tau_{\mathrm{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)$, is made as a cross-check. The resulting measurements, with only statistical uncertainties, are found to be

$$A_\Gamma^{\mathrm{K}\pi,\mathit{eff}} = (-0.9 \pm 2.2) \times 10^{-3}, \tag{5.24a}$$

$$A_\Gamma = (-5.9 \pm 5.9) \times 10^{-3}, \tag{5.24b}$$

$$y_{CP} = (5.5 \pm 6.3) \times 10^{-3}. \tag{5.24c}$$

The stability of these measurements and their systematic uncertainties are evaluated in chapter 6.

Chapter 6

Stability Verification and Evaluation of Systematic Uncertainties for y_{CP} and A_Γ

Chapter 5 presented the data and specific methods used to measure y_{CP} and A_Γ. A measurement of $A_\Gamma^{\mathrm{K}\pi,eff}$, defined analogously to A_Γ but using $\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ and $\tau_{\text{eff}}(\bar{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)$, is made as a cross-check. The resulting measurements, with only statistical uncertainties, are shown in equations 5.23a to 5.23c. Before the physical meaning of these results can be discussed they must be shown to be reliable and any systematic effects on their values evaluated.

Cross-checks of these results, performed by splitting the data according to various different criteria, are shown in section 6.1. Verification studies of the fit method itself, made using toy Monte Carlo simulation, are then shown in section 6.2. Finally, the systematic uncertainties on y_{CP} and A_Γ that may arise from various aspects of the event reconstruction and fit methodology are evaluated in section 6.3. The systematic uncertainties on $A_\Gamma^{\mathrm{K}\pi,eff}$ are also calculated, to ensure that they are compatible with those found for y_{CP} and A_Γ. The final results are then shown, and their physical interpretation discussed.

6.1 Measurement Cross-Checks

The measurement of $A_\Gamma^{\mathrm{K}\pi,eff}$, defined as

$$A_\Gamma^{\mathrm{K}\pi,eff} = \frac{\tau_{\text{eff}}(\bar{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-) - \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)}{\tau_{\text{eff}}(\bar{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-) + \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)}, \tag{6.1}$$

should be consistent with zero, as both $\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ and $\tau_{\text{eff}}(\bar{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)$ should yield the average lifetime of the D^0 mass eigenstates. The result shown in equation 5.23a, is very much consistent with this. As this exploits the high statistics of the $\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^\mp\pi^\pm$ channel this

alone provides a good deal of confidence in the measurements of A_Γ and y_{CP}. Additionally, the untagged lifetimes, shown in table 5.12, are equal to the average of the lifetimes for D^0 and $\overline{\mathrm{D}}^0$, as they should be. Finally, the untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ lifetime is nicely consistent with the world average value of [7]

$$\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+) = 410.1 \pm 1.4\,\text{fs}. \tag{6.2}$$

Thus, the basic measurements imply that the results are reliable.

To ensure that the fit results do not exhibit any dependence on any properties of the candidates or the running conditions, the data are split into bins depending on several different parameters. $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, A_Γ and y_{CP} are then measured on each individual dataset. The parameters by which the data are split are: the running period, to examine any dependence on the running conditions of LHCb; the D^0 momentum, to examine any dependence on the momentum scale; the D^0 p_T, for similar reasons; and the number of PVs in the event, which varies the level of lifetime biasing effects, as discussed in section 4.3.2.1. The χ^2-probability of the variations of the fitted values between the datasets is determined to evaluate their statistical significance. This is calculated with respect to the fit values obtained on the full datasets, using the uncorrelated statistical uncertainties. The uncorrelated uncertainties are calculated for the lifetimes using

$$\sigma^2_{\tau,X,\text{uncorr.}} = \sigma^2_{X,\text{bin}} + \sigma^2_{X,\text{full}} - 2\sqrt{\rho_X}\sigma_{X,\text{bin}}\sigma_{X,\text{full}}, \tag{6.3}$$

for A_Γ and $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ using

$$\begin{aligned}
\sigma^2_{A_\Gamma,\text{uncorr.}} = &\left(\frac{2\tau_{\overline{\mathrm{D}}^0,\text{bin}}}{(\tau_{\overline{\mathrm{D}}^0,\text{bin}}+\tau_{\mathrm{D}^0,\text{bin}})^2}\right)^2\sigma^2_{\mathrm{D}^0,\text{bin}} + \left(\frac{2\tau_{\mathrm{D}^0,\text{bin}}}{(\tau_{\overline{\mathrm{D}}^0,\text{bin}}+\tau_{\mathrm{D}^0,\text{bin}})^2}\right)^2\sigma^2_{\overline{\mathrm{D}}^0,\text{bin}} \\
&+ \left(\frac{2\tau_{\overline{\mathrm{D}}^0,\text{full}}}{(\tau_{\overline{\mathrm{D}}^0,\text{full}}+\tau_{\mathrm{D}^0,\text{full}})^2}\right)^2\sigma^2_{\mathrm{D}^0,\text{full}} + \left(\frac{2\tau_{\mathrm{D}^0,\text{full}}}{(\tau_{\overline{\mathrm{D}}^0,\text{full}}+\tau_{\mathrm{D}^0,\text{full}})^2}\right)^2\sigma^2_{\overline{\mathrm{D}}^0,\text{full}} \\
&- 2\sqrt{\rho_{\mathrm{D}^0}}\sigma_{\mathrm{D}^0,\text{bin}}\sigma_{\mathrm{D}^0,\text{full}}\frac{2\tau_{\overline{\mathrm{D}}^0,\text{bin}}}{(\tau_{\overline{\mathrm{D}}^0,\text{bin}}+\tau_{\mathrm{D}^0,\text{bin}})^2}\frac{2\tau_{\overline{\mathrm{D}}^0,\text{full}}}{(\tau_{\overline{\mathrm{D}}^0,\text{full}}+\tau_{\mathrm{D}^0,\text{full}})^2} \\
&- 2\sqrt{\rho_{\overline{\mathrm{D}}^0}}\sigma_{\overline{\mathrm{D}}^0,\text{bin}}\sigma_{\overline{\mathrm{D}}^0,\text{full}}\frac{2\tau_{\mathrm{D}^0,\text{bin}}}{(\tau_{\overline{\mathrm{D}}^0,\text{bin}}+\tau_{\mathrm{D}^0,\text{bin}})^2}\frac{2\tau_{\mathrm{D}^0,\text{full}}}{(\tau_{\overline{\mathrm{D}}^0,\text{full}}+\tau_{\mathrm{D}^0,\text{full}})^2},
\end{aligned} \tag{6.4}$$

and for y_{CP} using

$$\begin{aligned}
\sigma^2_{y_{CP},\text{uncorr.}} = &\left(\frac{1}{\tau_{KK,\text{bin}}}\right)^2\sigma^2_{K\pi,\text{bin}} + \left(\frac{\tau_{K\pi,\text{bin}}}{\tau^2_{KK,\text{bin}}}\right)^2\sigma^2_{KK,\text{bin}} \\
&+ \left(\frac{1}{\tau_{KK,\text{full}}}\right)^2\sigma^2_{K\pi,\text{full}} + \left(\frac{\tau_{K\pi,\text{full}}}{\tau^2_{KK,\text{full}}}\right)^2\sigma^2_{KK,\text{full}} \\
&- 2\sqrt{\rho_{K\pi}}\sigma_{K\pi,\text{bin}}\sigma_{K\pi,\text{full}}\frac{1}{\tau_{KK,\text{bin}}}\frac{1}{\tau_{KK,\text{full}}} \\
&- 2\sqrt{\rho_{KK}}\sigma_{KK,\text{bin}}\sigma_{KK,\text{full}}\frac{\tau_{K\pi,\text{bin}}}{\tau^2_{KK,\text{bin}}}\frac{\tau_{K\pi,\text{full}}}{\tau^2_{KK,\text{full}}},
\end{aligned} \tag{6.5}$$

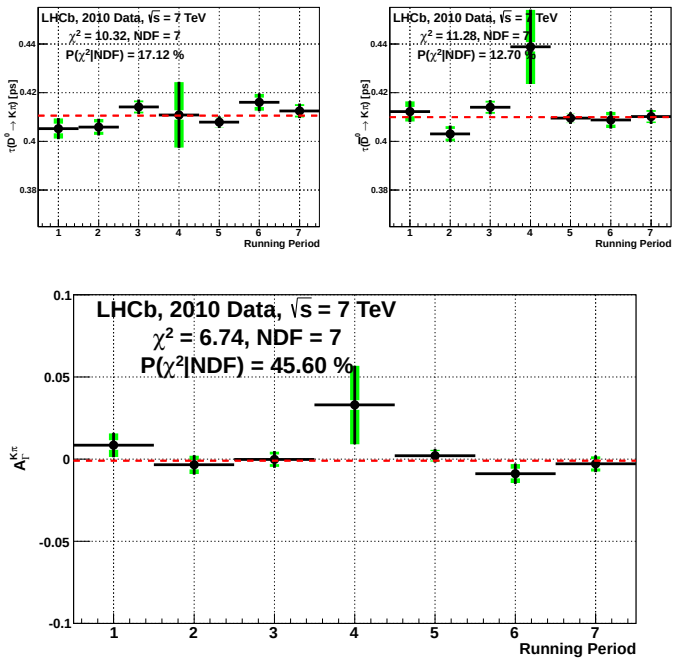
Bin No.	Run Range	N. Untagged $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$	N. Untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$
1	$80200 - 80375$	3,134	18,821
2	$80376 - 80650$	3,740	30,472
3	$80651 - 80875$	6,315	41,342
4	$80876 - 81309$	219	1,883
5	$81310 - 81375$	7,756	63,196
6	$81401 - 81475$	3,918	28,395
7	$81476 - 82000$	5,275	41,115

Table 6.1: The run ranges and the corresponding dataset sizes for the results shown in figures 6.1 and 6.2. The ranges are chosen to select data taken with different trigger configurations.

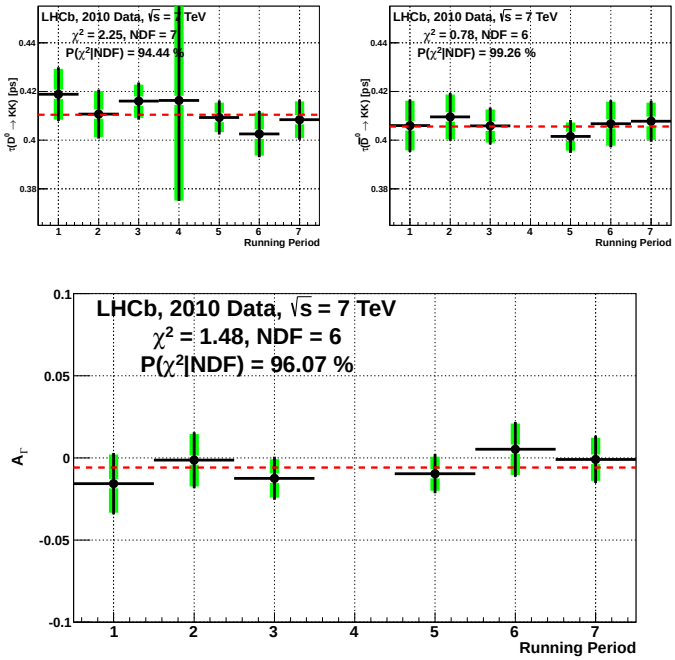
where $\tau_{X,\mathrm{bin}}$ is the measured lifetime in the given data bin for decay channel X, $\sigma_{X,\mathrm{bin}}$ its statistical uncertainty, $\tau_{X,\mathrm{full}}$ and $\sigma_{X,\mathrm{full}}$ the same for the full dataset, and ρ_X the fraction of the full dataset for the given decay channel that lies in the given bin. On occasion the datasets are too small after splitting the data for the fits to converge accurately, in which case the results are omitted from the comparisons.

The values obtained when splitting by running period are shown in figure 6.1 for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ and A_Γ, and figure 6.2 for y_{CP}. The run ranges chosen correspond to periods in which different trigger configurations were used. The number of candidates in each dataset for each running period is shown in table 6.1. The values of each measurement for each running period, as well as the effective lifetimes that contribute to them, agree within their statistical errors. The χ^2-probabilities also indicate that the variations between bins are consistent with statistical variations. The higher statistics of the untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ channel also allows examination of the pull distribution for $\tau_{\mathrm{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ using the results from each running period. The pulls are calculated with respect to the value obtained using the full untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ sample. The uncorrelated statistical uncertainties are used to calculate the pull. Their distribution is shown in figure 6.3, fitted with a single Gaussian. Although the uncertainties on the mean and σ are large they are clearly consistent with zero and one respectively. This also confirms that the measured effective lifetimes and their statistical uncertainties are determined correctly.

The results of splitting the data by D^0 p are shown for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ and A_Γ in figure 6.4 and for y_{CP} in figure 6.5. Splitting by D^0 p_T gives the results shown in figures 6.6 and 6.7. One might infer some weak dependence of $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ and A_Γ on p from these figures. However, these values are still consistent within their statistical errors, and the χ^2-probabilities are



(a)



(b)

Figure 6.1: (a) The fitted values of $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)$ (top right), and the resulting values of $A_\Gamma^{\mathrm{K}\pi,eff}$ (bottom) in different running periods. (b) The same for $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$, $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-)$ and A_Γ. The run ranges and statistics corresponding to each bin are shown in table 6.1. There are too few $\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-$ candidates in bin 4 for the fit to converge, so this is omitted from the comparison. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.

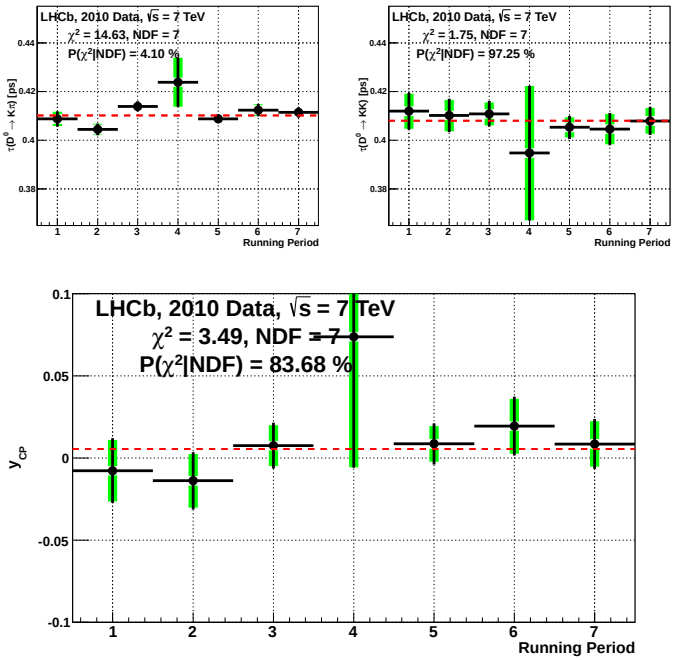


Figure 6.2: Fitted values of untagged $\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and untagged $\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$ (top right), and the resulting values of y_{CP} (bottom) in different running periods. The run ranges and statistics corresponding to each bin are shown in table 6.1. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.

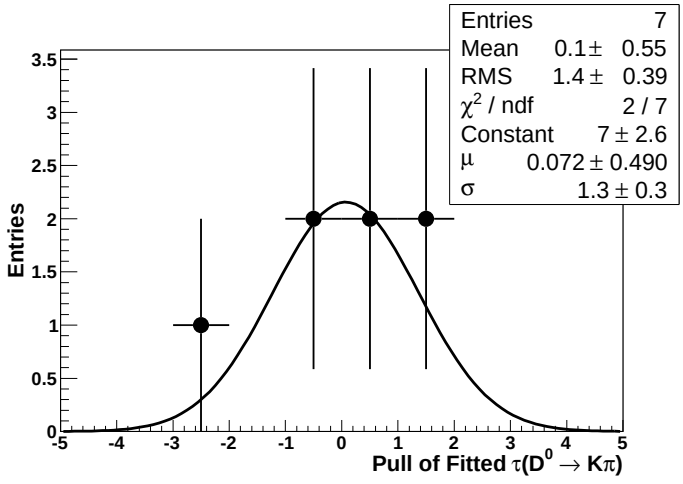
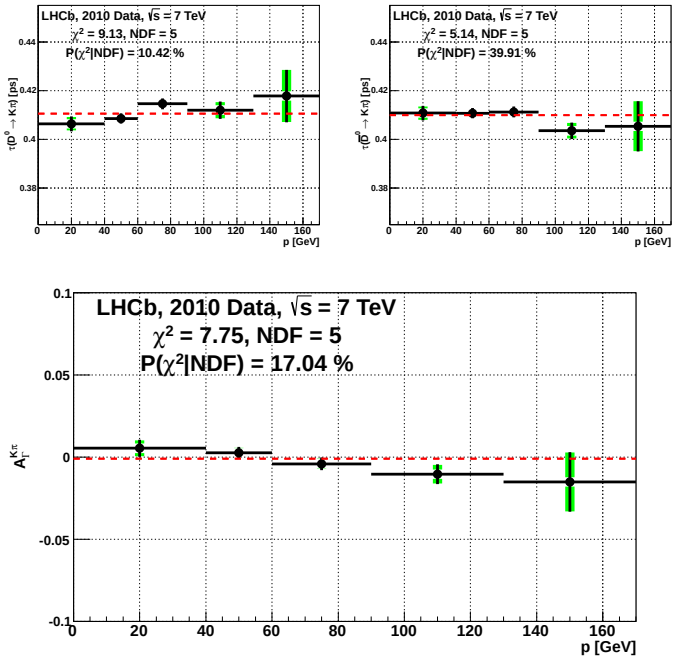
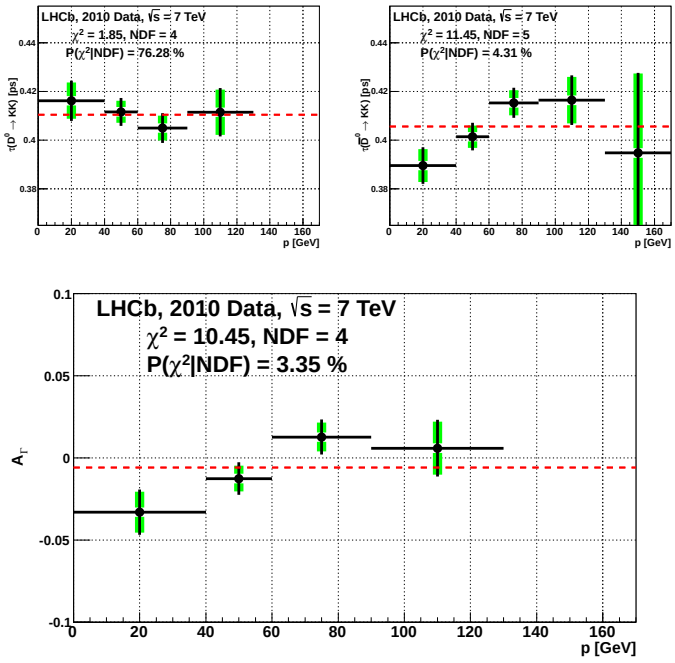


Figure 6.3: The pull distribution of the values of untagged $\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ when splitting the data by running period, taken with respect to the value obtained on the full dataset.



(a)



(b)

Figure 6.4: (a) The fitted values of $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)$ (top right), and the resulting values of $A_\Gamma^{\mathrm{K}\pi,eff}$ (bottom) in different bins of D^0 momentum. (b) The same for $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$, $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-)$ and A_Γ. The bin ranges are chosen to give roughly the same statistics in each bin. There are too few $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ candidates in the highest p bin for the fit to converge, so it is omitted from the comparison. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.

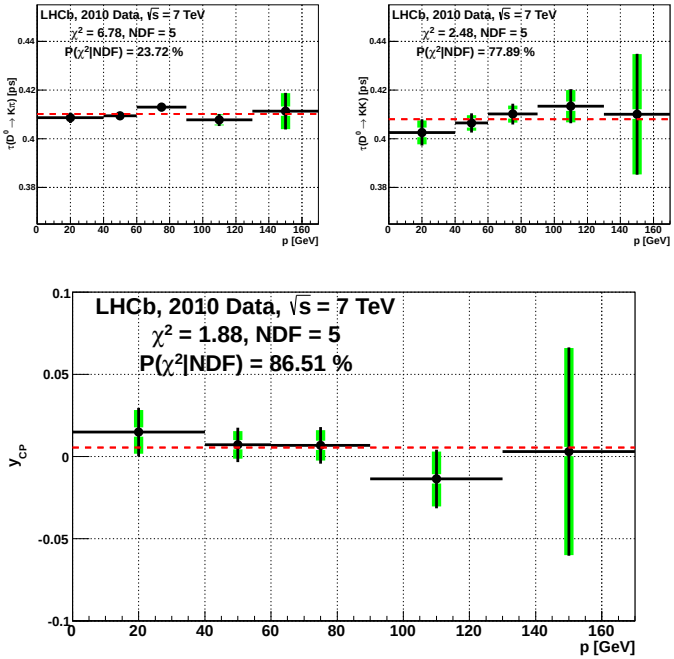
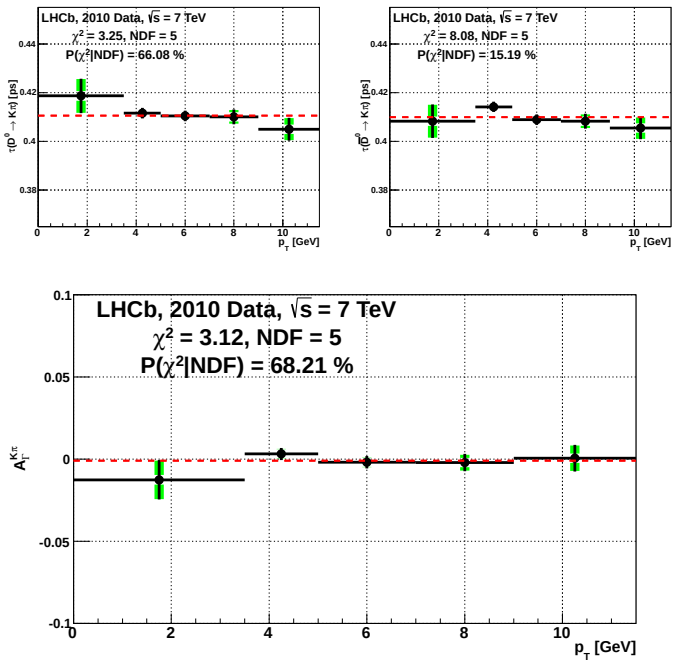
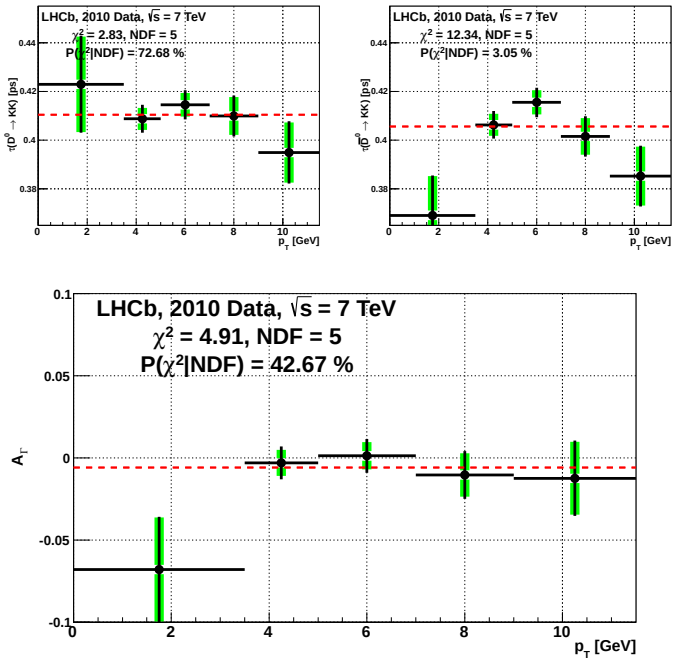


Figure 6.5: Fitted values of untagged $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and untagged $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$ (top right), and the resulting values of y_{CP} (bottom) in different bins of D^0 momentum. The bin ranges are chosen to give roughly the same statistics in each bin. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.



(a)



(b)

Figure 6.6: (a) The fitted values of $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)$ (top right), and the resulting values of $A_\Gamma^{\mathrm{K}\pi,eff}$ (bottom) in different bins of D^0 p_T. (b) The same for $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$, $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-)$ and A_Γ. The bin ranges are chosen to give roughly the same number of candidates in each bin. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.

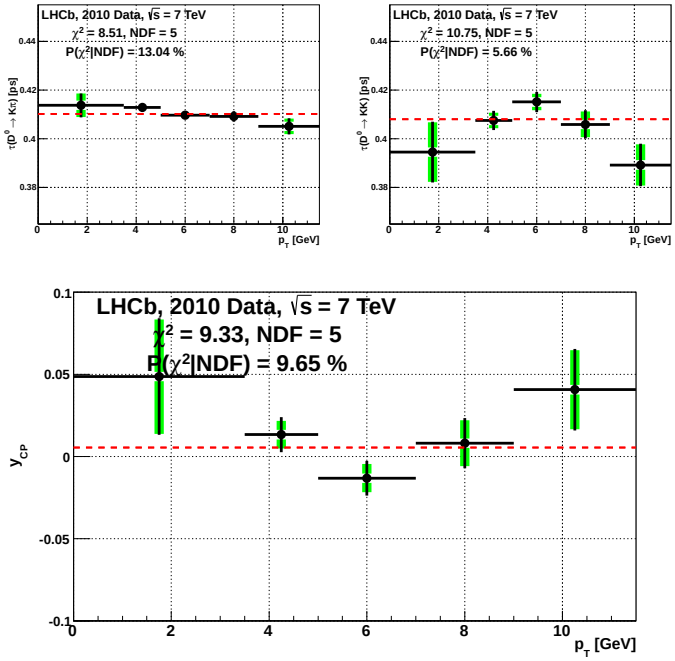


Figure 6.7: Fitted values of untagged $\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and untagged $\tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$ (top right), and the resulting values of y_{CP} (bottom) in different bins of D^0 p_T. The bin ranges are chosen to give roughly the same number of candidates in each bin. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.

consistent with statistical variations. Further, as the trends for $A_\Gamma^{\mathrm{K}\pi,eff}$ and A_Γ are in the opposite direction, and no similar dependence is seen when splitting by p_T, one can safely conclude that these apparent trends are simply due to statistical fluctuations. The values of y_{CP} obtained in each bin are also consistent with each other.

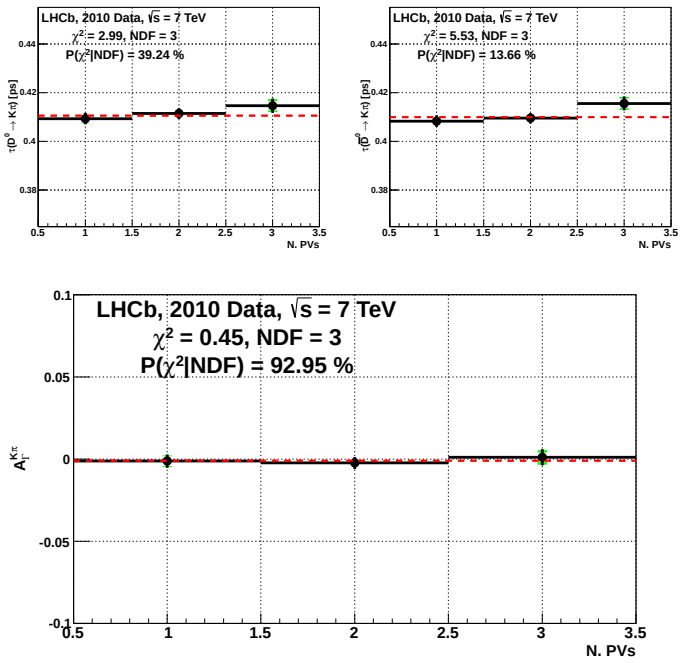
Finally, splitting the data by the number of PVs in each event gives the results shown in figures 6.8 and 6.9. As was discussed in section 4.3.2.1 higher PV multiplicity results in a greater number of turning points in the per-candidate acceptance functions determined by the swimming algorithm. Thus, examining candidates from events with different numbers of PVs may reveal any issues in accurately determining acceptance functions with varying numbers of turning points. The results show no significant dependence on the PV multiplicity, and so multiple turning point acceptance functions appear to be determined accurately.

These cross-checks thus show that the measurements of $A_\Gamma^{\mathrm{K}\pi,eff}$, A_Γ and y_{CP} are stable and exhibit no significant dependence on running period, D^0 p and p_T, and PV multiplicity.

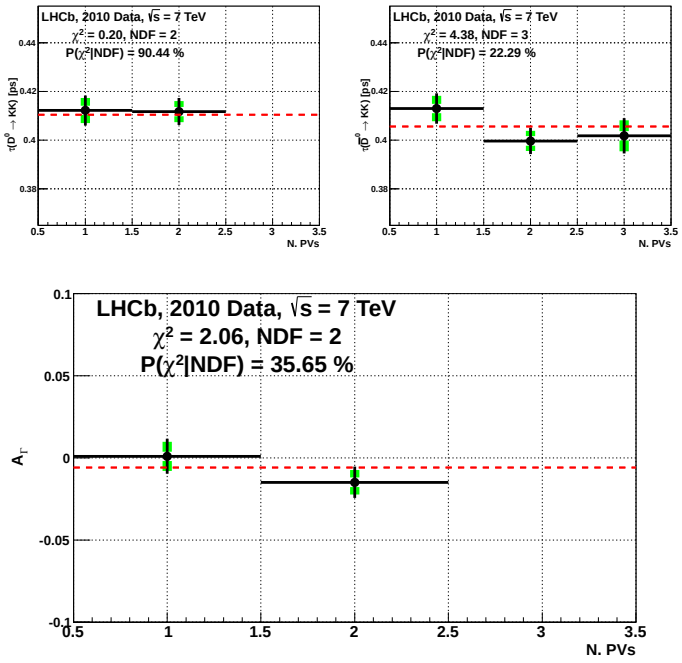
6.2 Verification Using Toy Monte Carlo Data

There are various ways in which a systematic bias can occur in the effective lifetimes determined by the method laid out in section 5.3.1. The cross-checks detailed in section 6.1 are insensitive to any such biases as they would affect all measurements in the same manner, regardless of how the data are divided up. To verify the accuracy of the fit method and ensure that any bias resulting from the fit method itself is negligible one must examine its results in many 'pseudo experiments' using Monte Carlo simulated 'toy data'. In such toy data the variables of interest for each candidate, such as the reconstructed $m(\mathrm{D}^0)$ or proper decay time, are generated using specific, known PDFs. Resolution effects are modelled by adding variables following Gaussian distributions to the generated variable of interest. Toy data is preferable over the full GEANT based simulation described in section 2.2.8.2 for such method verification as it is much less CPU intensive to produce, and so can easily be generated in much larger quantities. Also, generating the toy data using the exact same PDFs as are used in the fit means that any observed bias originates from the fit method itself, rather than any reconstruction biases.

For the toy data studied here the parameters of the PDFs used to generate the variables of interest are, where possible, those obtained from the fits to the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ dataset. These results are shown in tables 5.6 (a) and (c) for the mass fits and table 5.11(b) for the proper-decay-time fit. Prompt D^0 candidates are generated using a double Gaussian PDF for the mass, an exponential PDF with a Gaussian resolution term for the proper decay time, and a double bifurcated Gaussian plus a single Gaussian, with means independent of



(a)



(b)

Figure 6.8: (a) The fitted values of $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\pi^-)$ (top right), and the resulting values of $A_\Gamma^{\mathrm{K}\pi,eff}$ (bottom) in different bins of PV multiplicity. (b) The same for $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$, $\tau_{\rm eff}(\overline{\mathrm{D}}^0 \to \mathrm{K}^+\mathrm{K}^-)$ and A_Γ. There are too few $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ candidates with 3 PVs for the fit to converge, and so this dataset is omitted from the comparison. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.

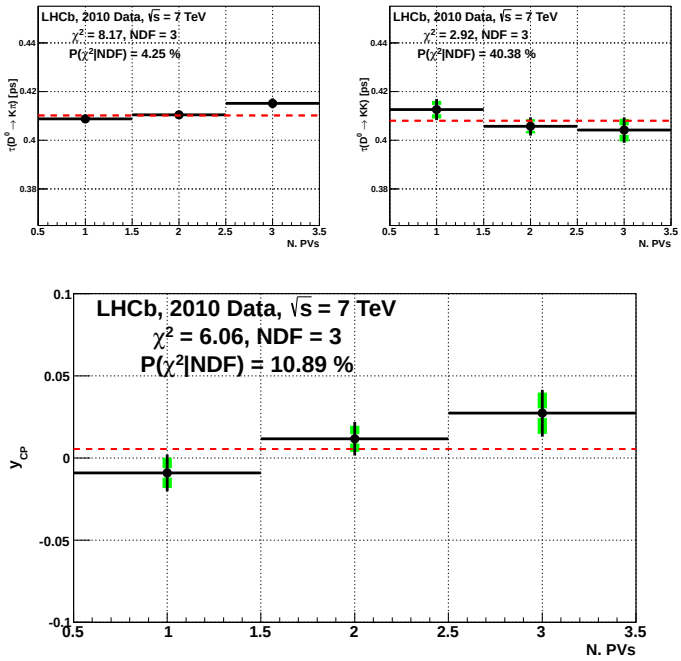
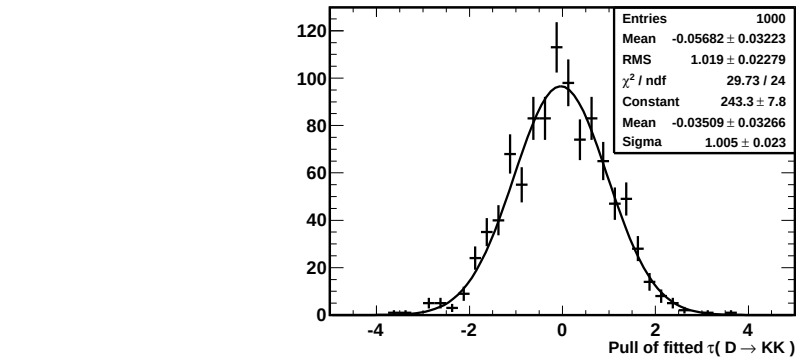


Figure 6.9: Fitted values of untagged $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^-\pi^+)$ (top left) and untagged $\tau_{\rm eff}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$ (top right), and the resulting values of y_{CP} (bottom) in different bins of PV multiplicity. The red dashed lines show the fitted values obtained using the whole dataset. The black error bars show the statistical uncertainties on each point, while the broad green error bars show the uncorrelated statistical uncertainties w.r.t. the red dashed line. The χ^2 and P-values shown are calculated w.r.t. the red dashed line using the uncorrelated uncertainties.

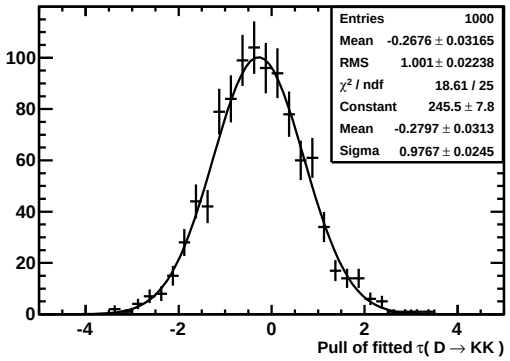
proper decay time, for the $\ln(\chi^2(IP_{\mathrm{D}^0}))$. Similarly, for secondary D^0 the mass is generated with the same PDF as prompt D^0, the proper decay time is generated using a sum of two exponential variables plus a Gaussian resolution term, and the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ using the same PDF as for prompt D^0 but with a proper-decay-time dependent mean of the form given by equation 5.11. Combinatorial background candidates are generated using a linear PDF for the mass, a single exponential PDF with Gaussian resolution for the proper decay time and a bifurcated Gaussian, with mean independent of proper decay time, for the $\ln(\chi^2(IP_{\mathrm{D}^0}))$. The parameters of the mass PDF for combinatorial backgrounds are taken from the results of the mass fit for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$, shown in table 5.6(a), while those of the proper decay time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDFs are taken from fits to the mass side-bands of the untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ dataset. Although the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distribution for combinatorial backgrounds is likely to have some proper-decay-time dependence, in reality it is not possible to determine this from the data. Generating the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ independent of proper decay time makes the combinatorial backgrounds more like signal than secondary D^0. Thus any bias originating from neglecting their PDFs in the proper-decay-time fit is likely to be slightly larger than in reality. Only 1 % of candidates in the mass side-bands are kept, to mirror the relative retention rate between the 'D2hh' and 'D2hh Wide Mass' trigger lines. No mis-tagged signal candidates are generated.

The first turning point value is generated with a double bifurcated Gaussian PDF. For prompt and secondary D^0 the parameters of this PDF are chosen to match the shape of the PDFs obtained using kernel density estimation from fits to data, like those shown in figure 5.9a. For combinatorial backgrounds the parameters are again obtained from fits to the mass side-bands of untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ data. The number of turning points is generated identically for all classes of candidate using a Gaussian PDF, the results of which are then rounded to positive integer values. The difference between turning points also follows the same distribution for all classes. It is generated using a single exponential PDF the mean of which is obtained from a fit to the untagged $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ dataset. The number of turning points and their values are generated before the proper decay time, and the generated proper decay time then required to lie within these acceptance intervals. No resolution is considered on the turning point values.

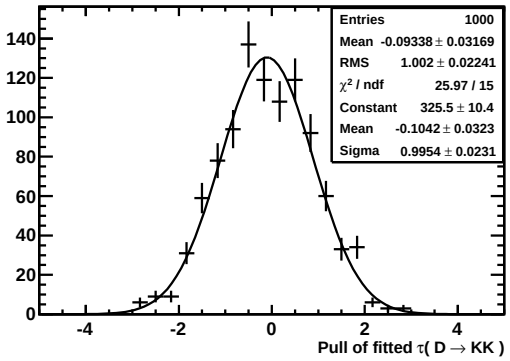
1000 such datasets of 20,000 candidates, of which 93 % are prompt D^0, 6 % secondary D^0, and 1 % combinatorial backgrounds, were generated. The full lifetime fit procedure, as detailed in section 5.3.1, was then applied to each. The pull of the fitted value of τ_{D^0}, defined as the fitted value minus the value used during the generation of the data, divided by the statistical error on the fitted value, is then plotted for all datasets. The pull distribution is shown in figure 6.10a, fitted with a Gaussian. The mean of this distribution is slightly



(a)



(b)



(c)

Figure 6.10: The pull distributions of the fitted values of τ_{D^0} obtained using 1000 toy datasets. In (a) the datasets consist of 20,000 candidates each with 93 % signal, 6 % secondary D^0, and 1 % combinatorial backgrounds. In (b) the datasets are identical but contain 100,000 candidates each. In (c) the datasets also contain 100,000 candidates each, but no combinatorial background is generated.

displaced from, but still consistent with, zero, and the σ is consistent with one. This shows that the fit method causes no significant bias to the lifetimes obtained and estimates the statistical errors correctly.

A further 1000 datasets were generated with 100,000 candidates each, and this process repeated. The corresponding pull distribution is shown in figure 6.10b. With the increased statistics a significant bias of 0.28 σ is apparent. However, this only equates to a bias of $\sim$0.4 fs on the lifetimes obtained. To verify the source of this bias 1000 datasets of 100,000 candidates each were generated with no combinatorial backgrounds. The mean of the resulting pull distribution, shown in figure 6.10c, is much closer to zero. This shows that the largest source of systematic bias to the fitted lifetimes results from neglecting combinatorial backgrounds in the proper-decay-time fit. A small bias may still remain, but as this corresponds to $\sim$0.1 fs it is negligible. The likely source of any remaining bias is inaccuracies in

reproducing the PDFs of the first turning point from data using kernel density estimation, as discussed in section 4.1.3.

Indeed, any consistent bias that displaces the measured lifetimes of D^0 and $\overline{\mathrm{D}}^0$ by the same amount, such as those shown in figure 6.10, will mostly cancel in the calculation of A_Γ. However, due to the different levels of background in the $\mathrm{K}^-\pi^+$ and $\mathrm{K}^+\mathrm{K}^-$ final states, a significant bias may result from neglecting combinatorial background when calculating y_{CP}. To examine what size of bias to expect, toy datasets were generated to mimic untagged $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ and $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ data. The $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ datasets have 230,000 candidates each, a signal lifetime of 410.1 fs and 1 % combinatorial background. The $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ datasets have 30,000 candidates each, a signal lifetime of 407.9 fs and 3 % combinatorial background. This roughly mimics the 2010 untagged datasets, with a generated y_{CP} of 5.5×10^{-3}. Fitting the proper-decay-time distribution in the mass side-bands of the 2010 datasets gives a mean lifetime of $\sim$420 $\pm$ 40 fs. To examine the effects of relatively extreme scenarios, two configurations were used to generate the toy data: one with the lifetime of the combinatorial background at 360 fs and one with the lifetime at 480 fs. 1000 datasets of $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ and $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ were then generated in both configurations, and the full fit procedure performed on each. The values of y_{CP} and their pull were then plotted for each configuration.

Figure 6.11a shows the fitted values of y_{CP} from the datasets with the combinatorial background generated with a lifetime of 360 fs. A small bias of $+1.3 \times 10^{-3}$ is observed. The corresponding pull distribution for y_{CP} is shown in figure 6.11b. The bias corresponds to $\sim$0.18 σ, but the statistical uncertainties are still estimated correctly. Figures 6.11c and 6.11d show the corresponding distributions for the datasets with a combinatorial background lifetime of 480 fs. Here a bias of -2.0×10^{-3}, corresponding to $\sim$0.28 σ, is observed. The statistical uncertainties are also estimated correctly.

The omission of combinatorial background is expected to be one of the main sources of systematic uncertainty in the measurements obtained. It is thus reassuring that in both these cases the biases observed are considerably smaller than the statistical uncertainties on y_{CP}. Further, as combinatorial background is generated in the toy data with $\ln(\chi^2(IP_{\mathrm{D}^0}))$ independent of proper decay time it is more similar to prompt D^0 than secondary D^0. In reality combinatorial background is expected to have some correlation between $\ln(\chi^2(IP_{\mathrm{D}^0}))$ and proper decay time, making it more like secondary D^0 than prompt D^0. The biases observed here can thus be taken as conservative upper estimates.

In both cases the statistical uncertainty on y_{CP} is 7.5×10^{-3}. Assuming a $1/\sqrt{N}$ dependence, and that the datasets are 100 % signal, one would expect the uncertainty to be 7.0×10^{-3}, using the size of the datasets after the cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$. The increase

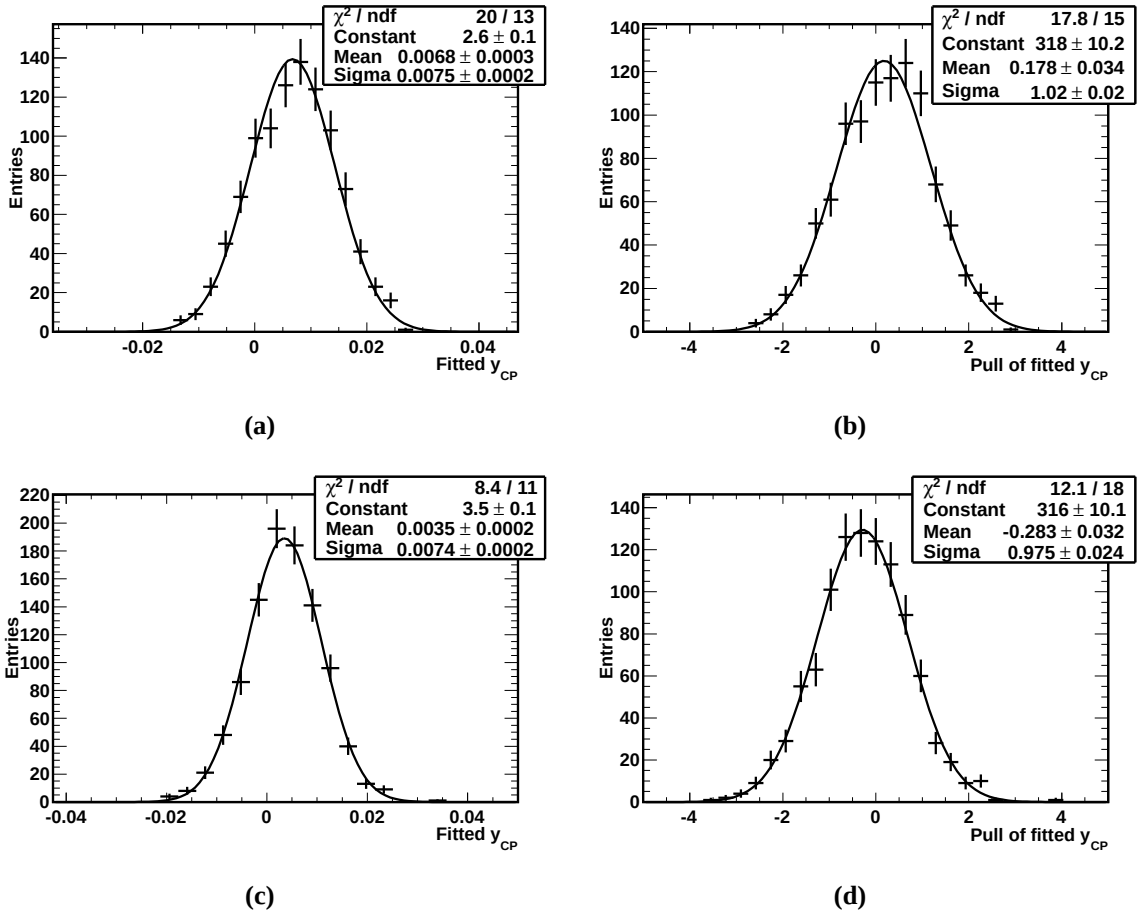


Figure 6.11: (a) The fitted values of y_{CP} from toy data when generating combinatorial background with a lifetime of 360 fs, and (b) the corresponding pull distribution. (c) and (d) show the same plots when generating combinatorial background with a mean lifetime of 480 fs.

in the observed uncertainty is likely due to the fact that the fit PDFs do not describe the data exactly when combinatorial background is neglected.

These studies thus show that the fit method described in section 5.3.1 and used to give the results shown in section 5.3.2 produces no significant biases to the measured parameters and estimates their statistical uncertainties accurately. What biases are observed can be attributed to neglecting combinatorial background in the fit PDF. Any resulting bias to the measured values of y_{CP} is expected to be considerably smaller than the statistical uncertainty achieved on its measured value.

6.3 Determination of Systematic Uncertainties

As mentioned previously, the cross-checks performed in section 6.1 demonstrate that the measurements of y_{CP} and A_Γ, presented with their statistical uncertainties in section 5.3.2, are stable and reliable, but do not give any indication as to the size of any systematic biases. Section 6.2 demonstrated that any bias resulting from the implementation of the fit itself is negligible, but also showed that a small bias is introduced as a result of neglecting combinatorial backgrounds in the fit for the effective lifetimes. As these results are found on ideal, toy data, any biases to the results obtained on the 2010 datasets are likely to be different. They also do not test for any biases resulting from the method of determining the per-candidate acceptance functions using the swimming algorithm, which is performed prior to and separately from the fit itself. To estimate the possible size of any such bias, various parameters within the fit are varied within a small range and the fits repeated. The systematic uncertainty resulting from each effect is taken as half of the total variation in the measured values of $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, A_Γ and y_{CP}. Variation of certain parameters change the size of the datasets used and so introduce some statistical variation into the fit results. In these cases the systematic uncertainty is still taken as the full range of variation in the results, in order to provide a conservative upper estimate of its value. The systematic uncertainties determined for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ are compared to those on A_Γ and y_{CP} to ensure compatibility.

6.3.1 Uncertainty on the VELO Length Scale

The proper decay time of a particle is determined using the distance between the PV and its decay vertex. Any bias to the relative positions of the VELO modules along the z-axis can thus result in a bias to the measured proper decay time. These positions were measured during the assembly of the VELO to an accuracy of 10 $\mu\mathrm{m}$, and again, once the VELO was sealed in the RF-box, using the track based alignment described in section 3.5. These two

measurements agree to an accuracy of 50 μm. Any potential bias the random displacement of an individual module may cause is reduced by the requirement that tracks have at least 9 hits in the VELO; the use of two tracks in calculating the position of the decay vertex; and the fact that the first hits on the tracks in any dataset are distributed across many different modules in both halves of the VELO. Consequently, the resulting bias to any lifetime measurement, such as those made here, is limited to be less than 0.04 % [99].

The track based alignment method is insensitive to a relative scaling of the z positions of the VELO modules. This is constrained only by the direct position measurements made during the assembly. At operational temperature the base plate of the VELO, to which the modules are attached, is maintained at 20 $^\circ C$ - slightly below room temperature. Thus, the base plate may have contracted slightly after the measurements made during assembly, which would introduce a scaling in the z positions of the modules. A conservative estimate of a 10 $^\circ$C temperature difference would correspond to a scaling of the z positions of the modules by $\sim 5 \times 10^{-5}$ [99]. This would translate into a similar level of bias to a lifetime measurement, which is negligible.

Uncertainties on the relative positions of the VELO and the downstream tracking stations can also introduce a lifetime bias. In particular, the position of the TT determined by track based alignment differs by 2 mm from the measurements made during its assembly. The worst case scenario for lifetime measurements would be if this was entirely due to the VELO z scale. This would introduce a lifetime bias of 0.1 %. Although the VELO z scale is known to a higher level of accuracy than this, as discussed above, this is conservatively assigned as an upper estimate to any resulting bias.

Hence, a systematic uncertainty of 0.1 % is assigned to the effective lifetime measurements made here. This corresponds to $\sim$0.4 fs for the D^0. As $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, A_Γ and y_{CP} all use ratios of lifetimes such a bias will cancel. The uncertainty from the VELO length scale is thus negligible.

6.3.2 Uncertainty on the Per-Candidate Acceptance Variables

As discussed in section 4.3.2.1 the swimming algorithm determines the positions of the turning points for the acceptance function for each candidate using an iterative refinement process. This results in a resolution on the proper decay time of the turning points of $\sim$0.5 fs ($\sim$4.6 μm). The worst case scenario would be if this process resulted in a consistent bias to the turning point values. To test this the turning point values for each candidate are displaced from their measured values and the fit re-run for each value of the displacement. A highly conservative bias of up to $\pm$ 3.4 fs (30 μm), corresponding to $\sim$6.5 σ, is applied. The results of these test are shown in figure 6.12 for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ and A_Γ and in figure 6.13 for y_{CP}.

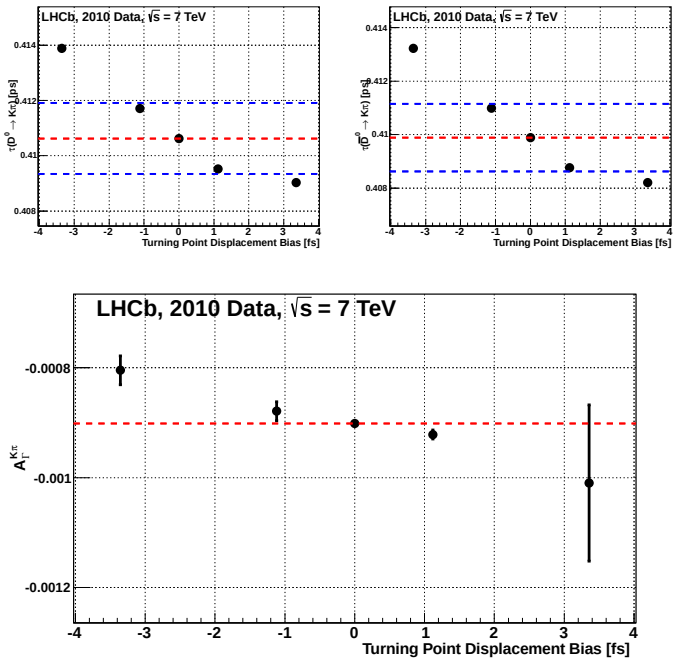
Systematic uncertainties of $\pm 0.10 \times 10^{-3}$, $\pm 0.17 \times 10^{-3}$, and $\pm 0.22 \times 10^{-3}$ respectively are thus assigned.

Proper-decay-time dependent differences between the proper-decay-time resolution in the trigger and offline reconstruction algorithms could also result in a consistent scaling of the turning point values. Thus, the turning point values are scaled by up to 1 ± 0.003 and the fit repeated for each scale value. A scale of 1 ± 0.001 is likely to be the worst case scenario, but the full range of 1 ± 0.003 is used to give a conservative estimate of any resulting bias. The fit results are shown in figures 6.14 and 6.15. An uncertainty of $\pm 0.049 \times 10^{-3}$, $\pm 0.15 \times 10^{-3}$, and $\pm 0.13 \times 10^{-3}$ is thus assigned to $A_\Gamma^{\mathrm{K}\pi,eff}$, A_Γ and y_{CP} respectively.

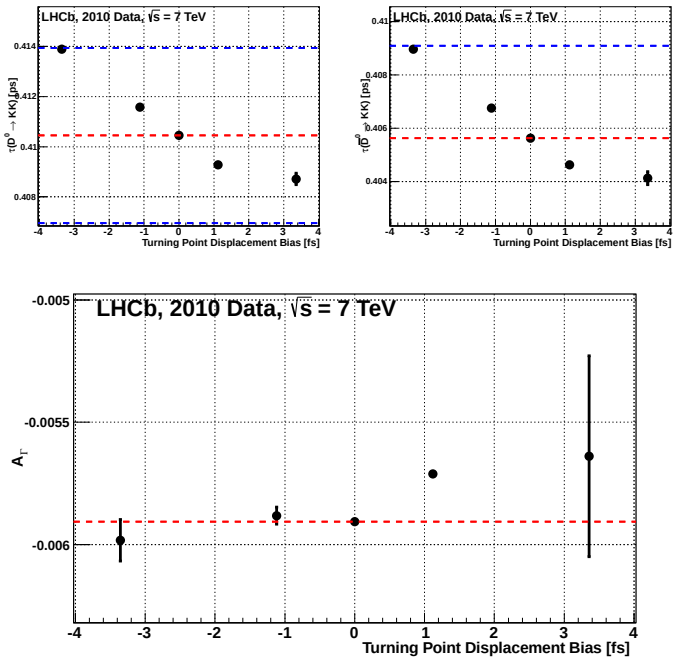
6.3.3 Uncertainty Due to Neglecting Combinatorial Backgrounds

The bias resulting from the omission of PDFs for combinatorial background in the fit is evaluated by varying the fraction of combinatorial background in the datasets. This is achieved by varying the size of the signal window in Δm. The default value is $\pm 2\,\mathrm{MeV}$, and so this is compared to using $\pm 1\,\mathrm{MeV}$ and $\pm 3\,\mathrm{MeV}$. This varies the fraction of combinatorial background in the $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ datasets between ∼0.46 % and ∼1.06 %. The relative variation for the $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ datasets should be of the same size. The results are shown in figures 6.16 and 6.17. These lead to systematic uncertainties of $\pm 1.5 \times 10^{-3}$, $\pm 1.3 \times 10^{-3}$, and $\pm 0.85 \times 10^{-3}$ for $A_\Gamma^{\mathrm{K}\pi,eff}$, A_Γ and y_{CP} respectively.

The uncertainty on y_{CP} is of a similar size to the bias observed in the studies on toy data, detailed in section 6.2. In addition to this studies were done whereby a fixed PDF for combinatorial background is added to the fit. The proper decay time is modelled as a single exponential. Four separate cases are considered: fixing the lifetime of the background to 360 fs or 480 fs, as in the toy studies; and having the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ PDF for the background the same as prompt or secondary D^0. The background fraction is fixed to 1 % for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and 3 % for $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$, and the fits repeated for each of the four configurations. The case of the background being like prompt D^0 in $\ln(\chi^2(IP_{\mathrm{D}^0}))$ results in a change to y_{CP} of -2.0×10^{-3} with the background lifetime at 360 fs, and $+4.4 \times 10^{-3}$ at 480 fs. Having the background like secondary D^0 in $\ln(\chi^2(IP_{\mathrm{D}^0}))$ results in a change of -0.1×10^{-3} and -0.2×10^{-3} for background lifetimes of 360 fs and 480 fs respectively. The change in this case is very small due to the suppression of the secondary-like component of the data in the final iteration of the fit. In reality any resulting bias is likely to lie between these two extremes. Thus, the systematic uncertainties determined by varying the signal window in Δm are taken as reasonable estimates.



(a)



(b)

Figure 6.12: The effects on the measured values of (a) $A_\Gamma^{\mathrm{K}\pi,eff}$, and (b) A_Γ, of biasing the turning point values by a small amount. The datasets in each bin are not fully correlated as biasing the turning points causes some candidates to have measured proper decay time outwith their acceptance intervals. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

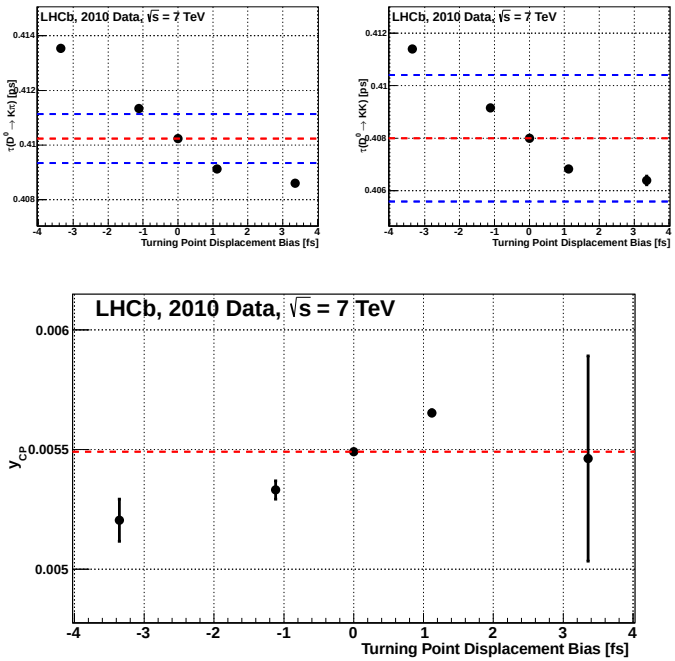
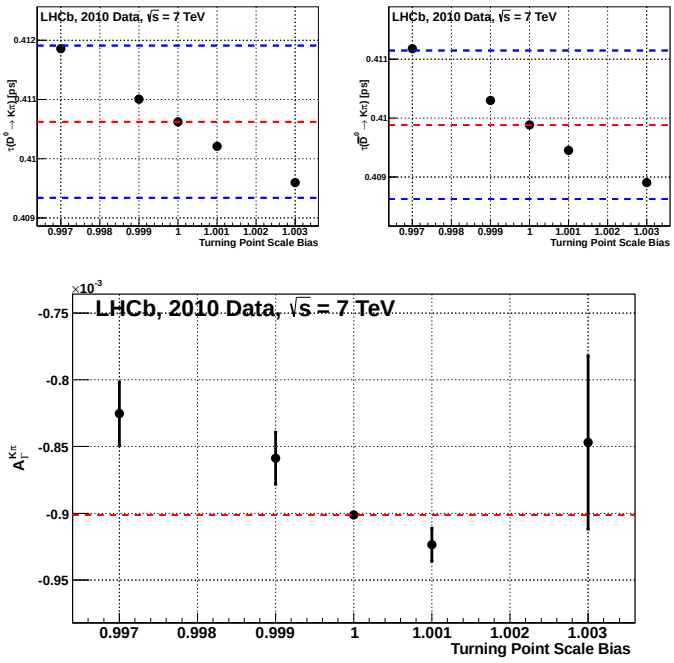
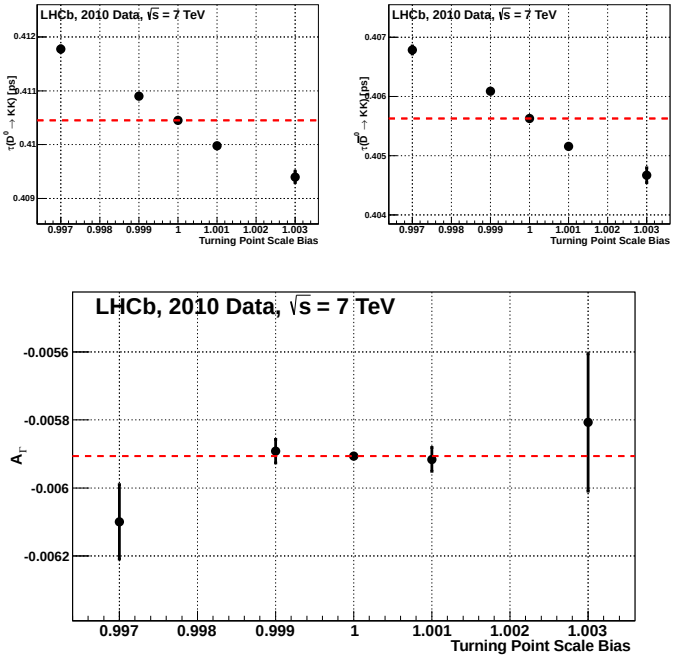


Figure 6.13: The effects on the measured value of y_{CP} of biasing the turning point values by a small amount. The datasets in each bin are not fully correlated as biasing the turning points causes some candidates to have measured proper decay time outwith their acceptance intervals. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.



(a)



(b)

Figure 6.14: The effects on the measured values of (a) $A_\Gamma^{\mathrm{K}\pi,eff}$, and (b) A_Γ, of scaling the turning point values by a small amount. The datasets in each bin are not fully correlated as biasing the turning points causes some candidates to have measured proper decay time outwith their acceptance intervals. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

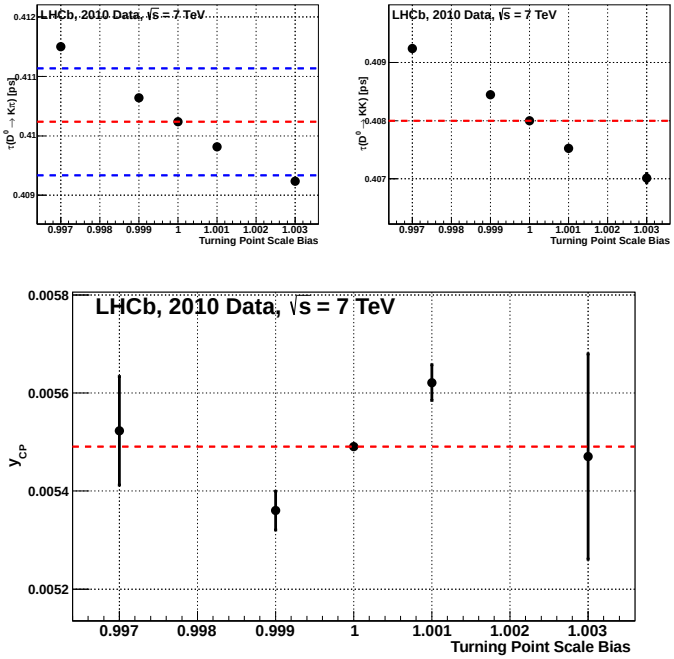
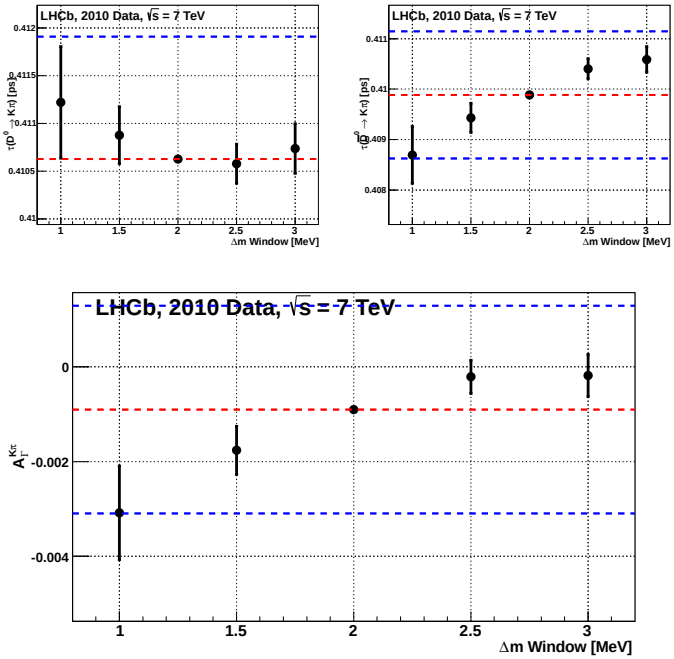
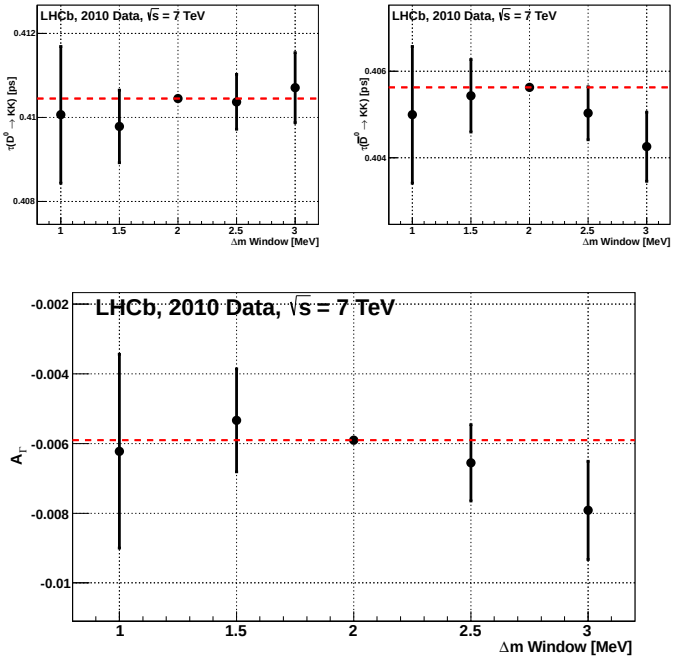


Figure 6.15: The effects on the measured value of y_{CP} of scaling the turning point values by a small amount. The datasets in each bin are not fully correlated as biasing the turning points causes some candidates to have measured proper decay time outwith their acceptance intervals. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.



(a)



(b)

Figure 6.16: The effect on (a) $A_\Gamma^{\mathrm{K}\pi,eff}$ and (b) A_Γ of varying the width of signal window in Δm from its nominal value of $\pm 2\,\mathrm{MeV}$. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

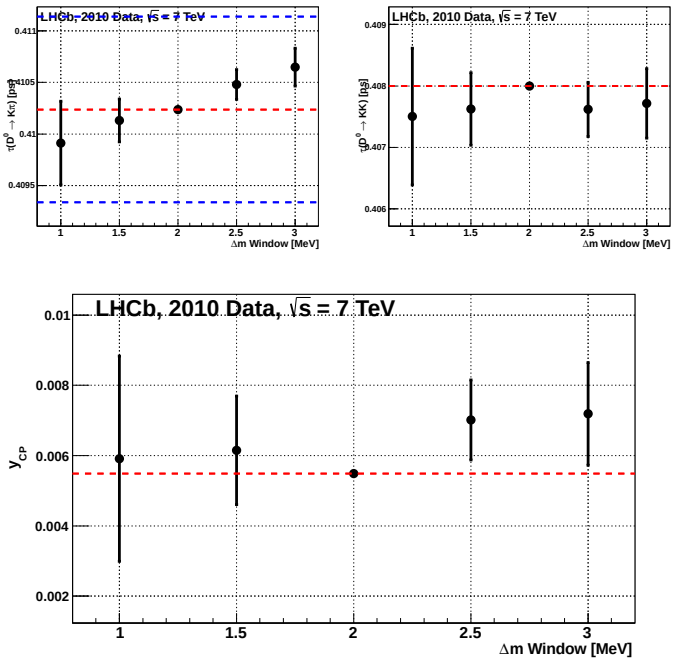


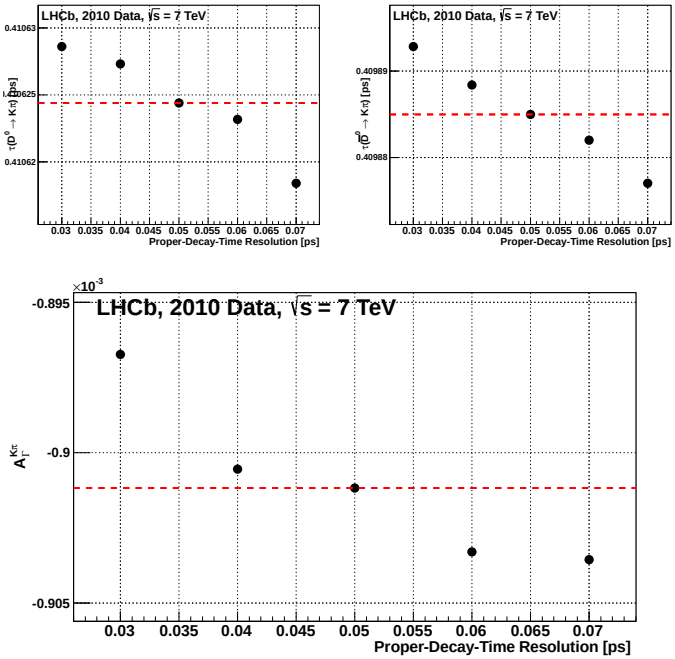
Figure 6.17: The effect on y_{CP} of varying the width of signal window in Δm from its nominal value of $\pm 2\,\mathrm{MeV}$. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

6.3.4 Uncertainty from the Parametrisation of Proper-Decay-Time Resolution

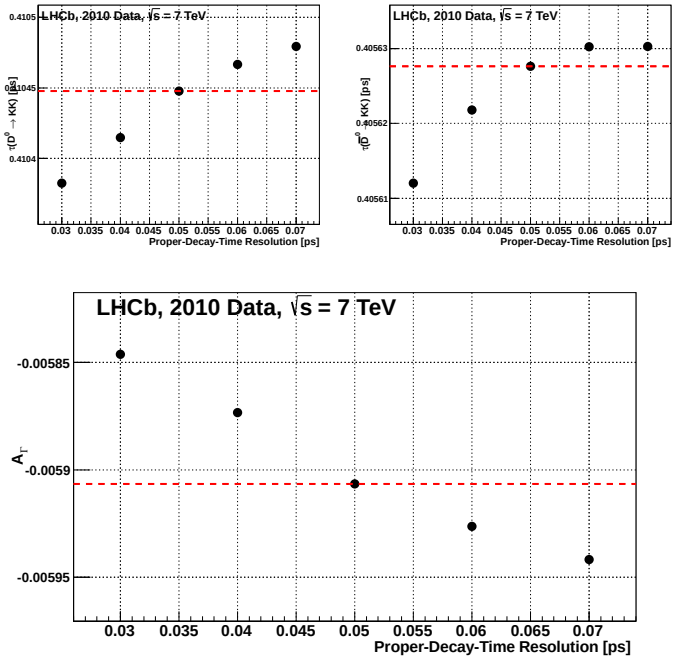
As discussed in section 5.3.1 the proper-decay-time resolution is modelled by a single Gaussian with width 50 fs. Any difference between this and the resolution function in reality may result in a systematic bias to the effective lifetimes. This parameter is also used as input to the calculation of the σ of the Gaussian kernel functions used to extract the PDFs of the first turning point, as described in section 5.3.1.3. To check the modelling of the proper-decay-time resolution and the accuracy of kernel density estimation in reproducing the distribition of TP_1 from the data the proper-decay-time resolution in the fit is varied between 30 fs and 70 fs. The results of this are shown in figures 6.18 and 6.19, and result in a systematic uncertainty of $\pm 0.0034 \times 10^{-3}$, $\pm 0.048 \times 10^{-3}$, and $\pm 0.056 \times 10^{-3}$ for $A_\Gamma^{\mathrm{K}\pi,eff}$, A_Γ, and y_{CP} respectively.

6.3.5 Uncertainty from the Boundaries of the Proper-Decay-Time Fit

As mentioned in section 5.1 the range of proper decay times in the fit is restricted to be between 0.25 ps and 6 ps. The lower limit is placed to avoid instabilities in the fit in regions of very low statistics, while the upper limit is used to exclude very long lived backgrounds.



(a)



(b)

Figure 6.18: The results for (a) $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$ and (b) A_Γ of varying the width of the proper-decay-time resolution function from its nominal value of 50 fs. The datasets in each bin are 100 % correlated and so no uncertainties are shown. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

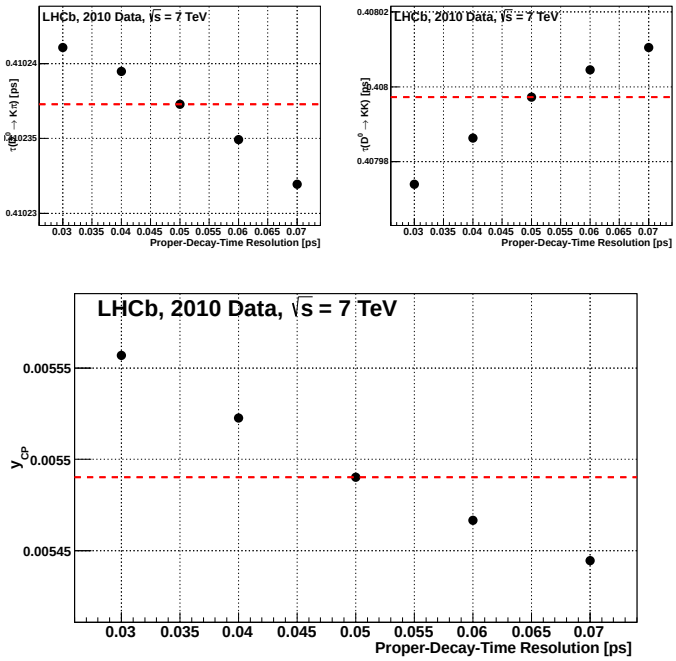


Figure 6.19: The results for y_{CP} of varying the width of the proper-decay-time resolution function from its nominal value of 50 fs. The datasets in each bin are 100 % correlated and so no uncertainties are shown. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

Varying these values thus allows evaluation of the accuracy of the fit model and the impact of such long lived backgrounds. This also examines the effects of the assumption that the difference between the first and second turning points follows the same distribution for prompt and secondary D^0. As discussed in section 5.3.1.7 this may not be strictly true in the presence of an upper limit on the proper decay time.

In figures 6.20 and 6.21 the lower bound is varied from 0.25 ps to 0.2 ps and 0.3 ps. This results in systematic uncertainties of $\pm 0.092 \times 10^{-3}$, $\pm 0.14 \times 10^{-3}$, and $\pm 0.75 \times 10^{-3}$ for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, A_Γ and y_{CP} respectively. In figures 6.22 and 6.23 the upper bound is varied from 6 ps to 5 ps and 8 ps. This results in systematic uncertainties of $\pm 0.073 \times 10^{-3}$, $\pm 0.21 \times 10^{-3}$, and $\pm 0.15 \times 10^{-3}$ for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, A_Γ and y_{CP} respectively.

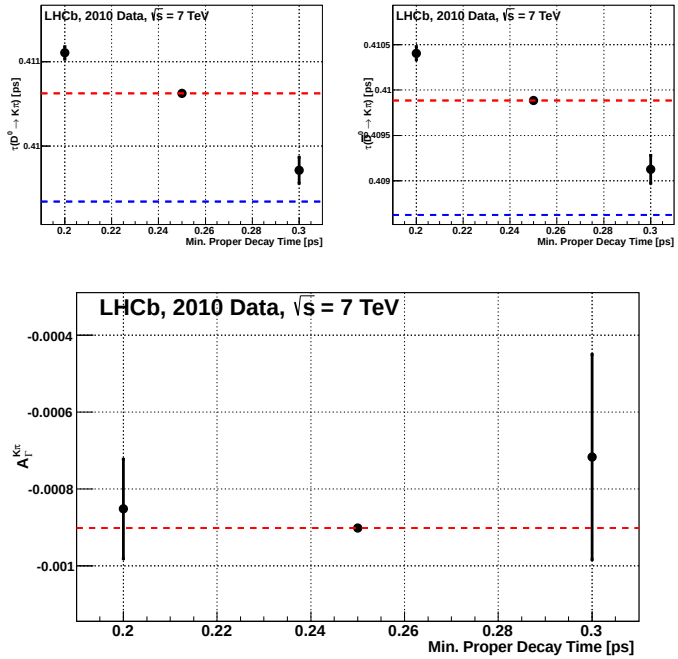
6.3.6 Uncertainty from the Parametrisation of Secondary D^0

Secondary D^0 and backgrounds with a similar topology are strongly suppressed in the final iteration of the proper-decay-time fit by the cut of $\ln(\chi^2(IP_{\mathrm{D}^0})) < 2$. To examine how well the remaining fraction of secondary D^0 is modelled in the fit this cut is varied between 1.5, which is just above the peak of the $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distribution for prompt signal, and 3.5. This varies the fraction of secondary D^0 determined to be in the final iteration of the fit between $\sim$0.4 % to $\sim$1.6 %. Any potential correlation between the values of the turning points and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ for secondary D^0, as discussed in section 5.3.1.7, will also vary with the value of this cut. As combinatorial backgrounds will tend to have larger $\chi^2(IP_{\mathrm{D}^0})$ this test also varies their relative fraction in the datasets. The results of this are shown in figures 6.24 and 6.25. These result in systematic uncertainties of $\pm 0.63 \times 10^{-3}$, $\pm 1.7 \times 10^{-3}$, and $\pm 3.9 \times 10^{-3}$ for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, A_Γ and y_{CP} respectively.

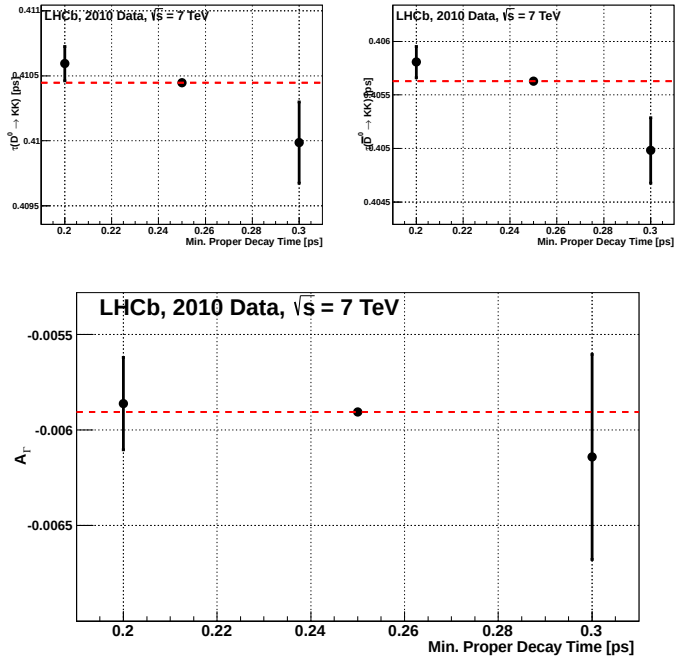
This variation is particularly large for y_{CP}. From figure 6.25 it is clear that this is due to the variation in $\tau_{\mathrm{eff}}(\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-)$ at low cut values. This is potentially due mostly to the contribution from combinatorial background, as the fraction of combinatorial background for $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ is much larger than for $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$. The combination of the uncertainty arising from this study and that from the variation of the Δm window thus likely over-estimates the effects of combinatorial background. Nonetheless, both values are conservatively included in the total systematic uncertainty.

6.3.7 Uncertainties Due to Varying Reconstruction Inefficiencies

As mentioned in section 4.3.2.1 the reconstruction efficiency of the D^0 can vary as a function of its transverse flight distance due to the assumption in the reconstruction algorithms that all tracks originate from the beam-line. The swimming algorithm is insensitive to such effects



(a)



(b)

Figure 6.20: The effects on (a) $A_\Gamma^{\mathrm{K}\pi,eff}$ and (b) A_Γ of varying the lower bound on proper decay time from its nominal value of 0.25 ps. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

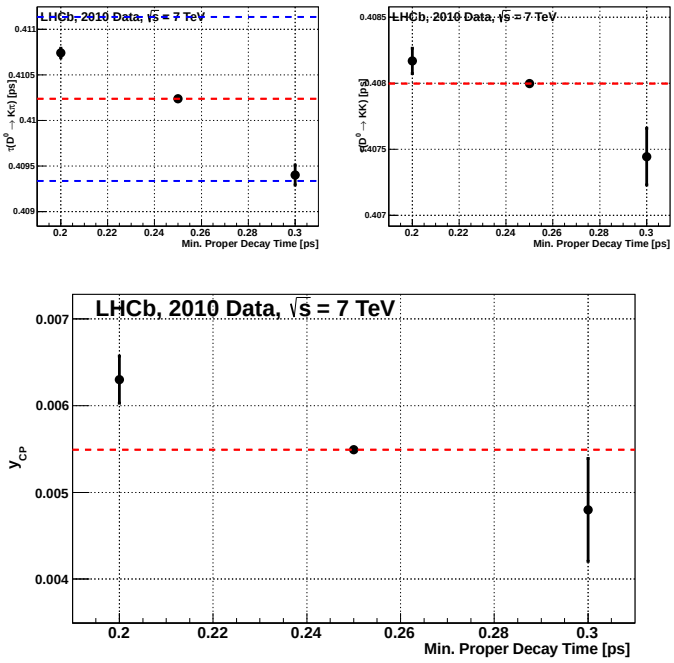
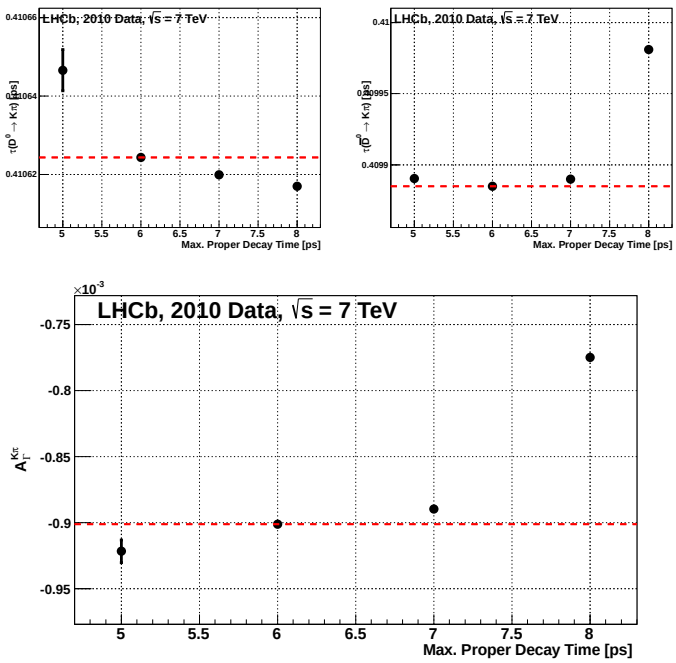


Figure 6.21: The effects on y_{CP} of varying the lower bound on proper decay time from its nominal value of 0.25 ps. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

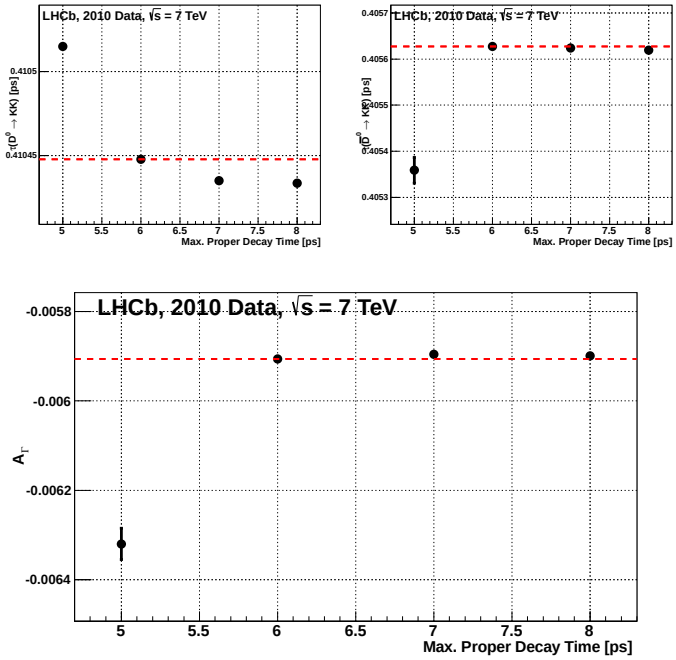
as the PVs are moved in order to change the D^0 proper decay time, rather than moving the decay vertex and the tracks made by the daughter particles. Thus any such effect would result in a bias to the measured lifetimes.

The existence of any reconstruction bias is evaluated using full Monte Carlo simulated data. The number of D^0 generated is compared to the number reconstructed, using the offline reconstruction, as a function of their proper decay time. The data are further divided up into bins of p and η to examine any geometric dependence on the reconstruction efficiency. A linear fit is performed to the efficiency distribution in each bin. Figure 6.26 shows the efficiency as a function of proper decay time for candidates with $50 < p\,[\mathrm{GeV}] < 70$. The gradient of the linear fit is $\beta = (2.5 \pm 5.0) \times 10^{-3}$, showing no significant variation in the efficiency as a function of proper decay time. Figure 6.27 shows the gradients of fits to the efficiency distributions in bins of η and p. No significant deviations from zero are observed. Thus, no reconstruction bias to the measured lifetimes is observed, and no systematic uncertainty is assigned.

Any similar bias resulting from the HLT reconstruction algorithms can be checked on real data by calculating the reconstruction efficiency with respect to the offline reconstruction. This is achieved using a sample of D^0 reconstructed from the minimum bias trigger



(a)



(b)

Figure 6.22: The effects on (a) $A_\Gamma^{\mathrm{K}\pi,eff}$ and (b) A_Γ of varying the upper bound on proper decay time from its nominal value of 6 ps. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

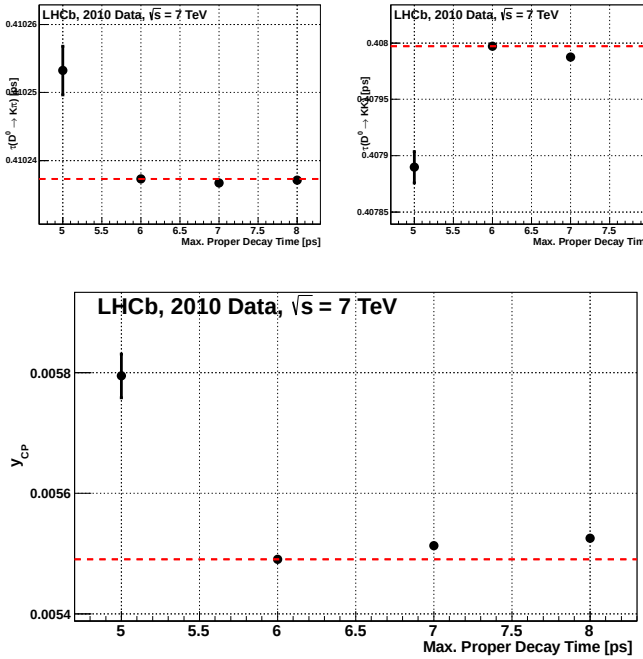


Figure 6.23: The effects on y_{CP} of varying the upper bound on proper decay time from its nominal value of 6 ps. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

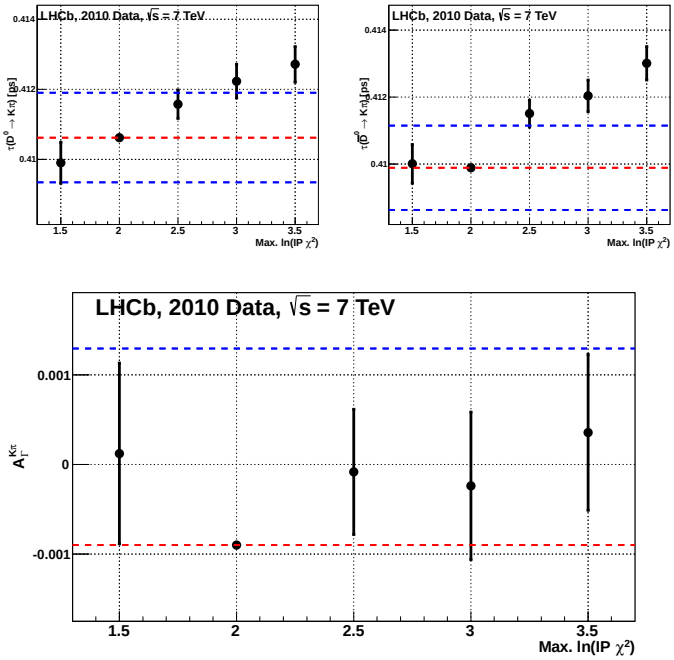
line, in which no lifetime biasing cuts are applied. The relative efficiency as a function of proper decay time is shown for HLT1 in figure 6.28a and HLT2 in figure 6.28b, fitted with a constant. The constant fit describes the data well, thus showing that there is no significant reconstruction bias. No systematic uncertainty for reconstruction biases in the HLT is thus applied.

6.3.8 Summary of Systematic Uncertainties and Final Results

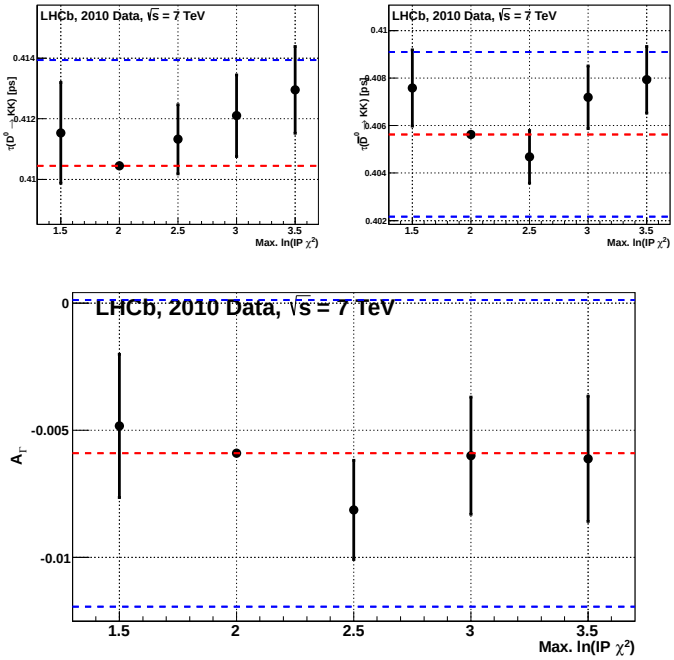
The systematic uncertainties assigned from each study detailed in the previous section for $A_\Gamma^{\mathrm{K}\pi,\mathit{eff}}$, A_Γ and y_{CP} are detailed in table 6.2. The contribution of each effect is assumed to be independent of the others, thus the total systematic uncertainty is calculated as the sum in quadrature of the uncertainties for each contributing effect. The dominant systematics arise from the parametrisation of secondary D^0 and neglecting combinatorial backgrounds in the fit. This gives the final result for the cross-check measurement

$$A_\Gamma^{\mathrm{K}\pi,\mathit{eff}} = (-0.9 \pm 2.2\,(\mathrm{stat.}) \pm 1.6\,(\mathrm{syst.})) \times 10^{-3}, \tag{6.6}$$

which is consistent with zero, as expected. The systematic uncertainties determined for this measurement are of a similar size to those found for A_Γ and y_{CP}. This shows that the method



(a)



(b)

Figure 6.24: The effects on (a) $A_\Gamma^{\mathrm{K}\pi,eff}$ and (b) A_Γ of varying the maximum cut on $\ln(\chi^2(IP_{\mathrm{D}^0}))$ in the final fit iteration from its nominal value of 2. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

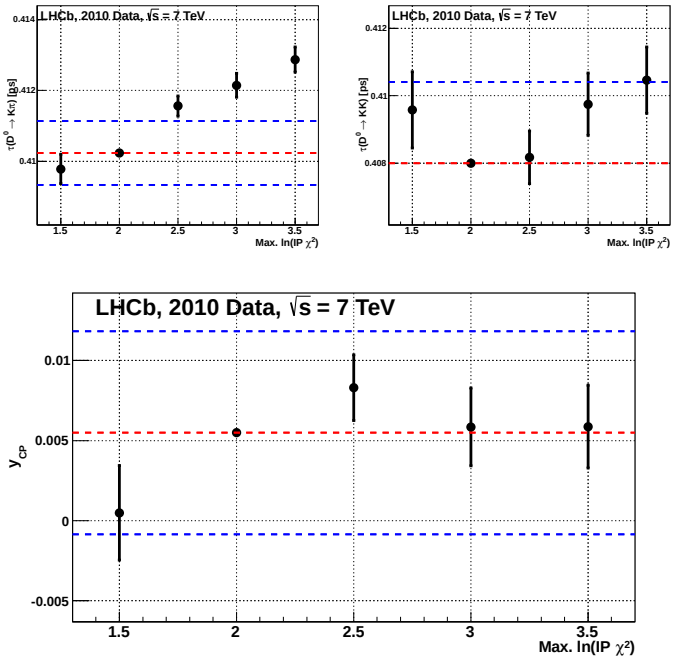


Figure 6.25: The effects on y_{CP} of varying the maximum cut on $\ln(\chi^2(IP_{\mathrm{D}^0}))$ in the final fit iteration from its nominal value of 2. The uncertainties shown are uncorrelated with respect to the nominal result. The red dashed line shows the nominal result. Blue dashed lines are drawn at the nominal result $\pm 1\times$ its statistical uncertainty, if this is in range.

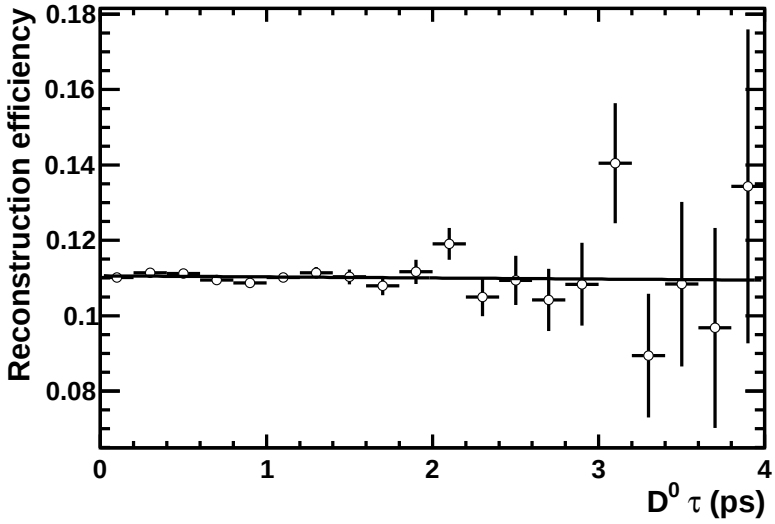
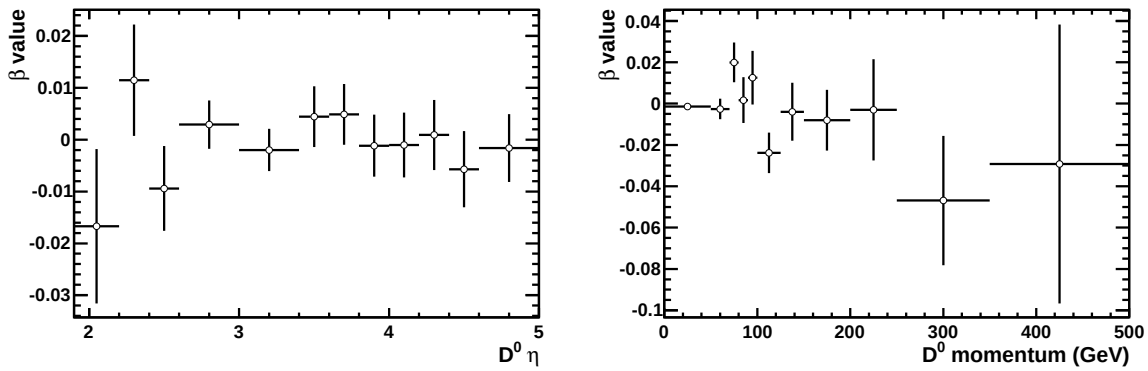
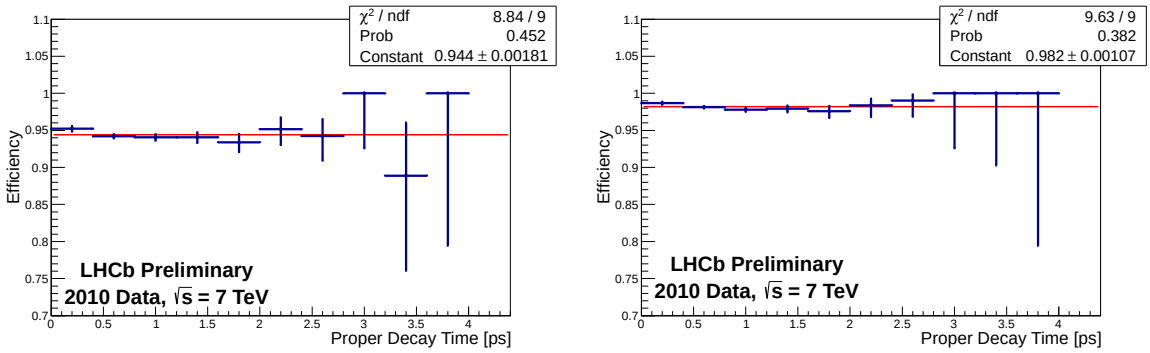


Figure 6.26: The reconstruction efficiency as a function of proper decay time from simulated data, for candidates with $50 < p\,[\mathrm{GeV}] < 70$. The linear fit has gradient $\beta = (2.5 \pm 5.0) \times 10^{-3}$. Reproduced from [87].



(a) **(b)**

Figure 6.27: The gradients, β, of linear fits to the reconstruction efficiency as a function of proper decay time in bins of η (a) and p (b), from simulated data. Reproduced from [87].



(a) **(b)**

Figure 6.28: The efficiency of the reconstruction in (a) HLT1 and (b) HLT2 with respect to the offline reconstruction, as a function of proper decay time, fitted with a constant.

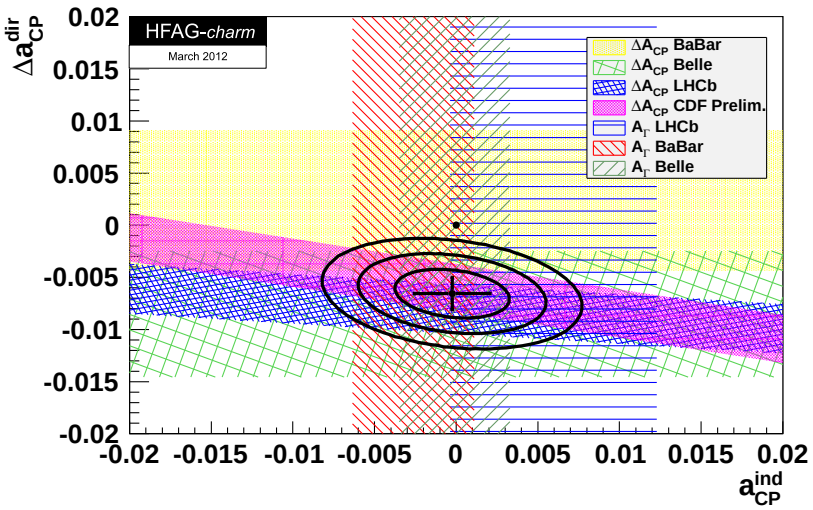


Figure 6.29: Combined world averages of direct CP-violation and indirect CP-violation in the D^0 system, including the measurement of A_Γ presented here. Calculated by the Heavy Flavour Averaging Group [3].

of evaluating systematic uncertainties is also reliable. The final measurements of A_Γ and y_{CP} are found to be

$$A_\Gamma = (-5.9 \pm 5.9\,(\text{stat.}) \pm 2.1\,(\text{syst.})) \times 10^{-3}, \tag{6.7a}$$

$$y_{CP} = (5.5 \pm 6.3\,(\text{stat.}) \pm 4.1\,(\text{syst.})) \times 10^{-3}. \tag{6.7b}$$

These results have been submitted for publication [33]. This is the first time these measurements have been performed at a hadron collider experiment. A_Γ is consistent with zero, and y_{CP} is consistent with the world average value of $y = (8.0 \pm 1.3) \times 10^{-3}$ [3]. Thus no indication of CP-violation is observed in these results. Indeed, this value of y_{CP} is also consistent with zero, and so shows no evidence for mixing in the D^0 system. These results are not yet competitive with the world best measurements made by the B factories, discussed in section 1.3, but are consistent with them. Nonetheless, they make an important contribution to the world average values. The average measurements of direct and indirect CP-violation in the D^0 sector including this measurement of A_Γ, combined by the Heavy Flavour Averaging Group [3], are shown in figure 6.29. The combined average is currently dominated by the measurement of $\Delta\mathcal{A}^{\mathrm{CP}}$ performed by LHCb [1], and sits above 3σ from the zero CP-violation hypothesis. Also included is a preliminary measurement of $\Delta\mathcal{A}^{\mathrm{CP}}$ by the CDF collaboration, which is in agreement with that of LHCb. The average value of A_Γ is still consistent with zero.

Systematic Effect	$A_\Gamma^{K\pi,eff}$ (10^{-3})	A_Γ (10^{-3})	y_{CP} (10^{-3})
VELO length scale	$< \pm 0.001$	$< \pm 0.001$	$< \pm 0.001$
Turning point bias	± 0.10	± 0.17	± 0.22
Turning point scaling	± 0.049	± 0.15	± 0.13
Combinatorial background (varying Δm window)	± 1.5	± 1.3	± 0.85
Proper time resolution	± 0.0034	± 0.048	± 0.056
Minimum proper time cut	± 0.092	± 0.14	± 0.75
Maximum proper time cut	± 0.073	± 0.21	± 0.15
Secondary D^0 (varying maximum $\ln(\chi^2(IP_{D^0}))$ cut)	± 0.6	± 1.7	± 3.9
Reconstruction bias	$< \pm 0.1$	$< \pm 0.1$	$< \pm 0.1$
Total systematic uncertainty	± 1.6	± 2.1	± 4.1
Statistical uncertainty	± 2.2	± 5.9	± 6.3
Measured value	-0.9	-5.9	5.5

Table 6.2: Summary of systematic uncertainties detailed in section 6.3.

6.4 Conclusions

This chapter presented verification of the stability of the measurements of y_{CP} and A_Γ presented in chapter 5, and evaluated the systematic uncertainties on their values. The same is done for the cross-check measurement of $A_\Gamma^{K\pi,eff}$ to ensure compatibility with the values obtained for y_{CP} and A_Γ.

Several cross-checks on the measurements of $A_\Gamma^{K\pi,eff}$, A_Γ and y_{CP}, were performed and the results shown in section 6.1. No significant dependence of the results on running period, D^0 p and p_T, and event PV multiplicity is observed, thus demonstrating that the results are stable. Section 6.2 described how the full fit was applied to toy Monte Carlo simulated data in order to evaluate any measurement bias resulting from the fit method. The only significant bias observed results from neglecting combinatorial backgrounds in the lifetime fit, though this bias is still much smaller than the statistical uncertainty on the measurements obtained.

Many possible sources of systematic bias were evaluated in section 6.3. The dominant systematic effects were found to be from neglecting combinatorial backgrounds in the lifetime fit, and varying the fraction of secondary D^0 in the datasets. The final value of and uncertainty on the cross-check measurement $A_\Gamma^{K\pi,eff}$ is found to be

$$A_\Gamma^{K\pi,eff} = (-0.9 \pm 2.2\,(\text{stat.}) \pm 1.6\,(\text{syst.})) \times 10^{-3}, \tag{6.8}$$

which is consistent with zero, as expected. The final results for y_{CP} and A_Γ are

$$A_\Gamma = \quad (-5.9 \pm 5.9\,(\text{stat.}) \pm 2.1\,(\text{syst.})) \times 10^{-3}, \tag{6.9a}$$

$$y_{CP} = \quad (5.5 \pm 6.3\,(\text{stat.}) \pm 4.1\,(\text{syst.})) \times 10^{-3}. \tag{6.9b}$$

These are the first measurements of these values at a hadron collider experiment, and have been submitted for publication [33]. A_Γ is consistent with zero, and y_{CP} is consistent with the world average value of $y = (8.0 \pm 1.3) \times 10^{-3}$ [3] and with zero. Thus no indication of CP-violation or mixing is observed in these results. Although the statistical uncertainties attained are not yet competitive with previous measurements made at B factories they make an important contribution to the world average.

Chapter 7

Conclusions and Outlook

7.1 Summary

This thesis presented measurements of the charm sector mixing and CP-violation parameters y_{CP} and A_Γ. These were first introduced in chapter 1 in the context of the mathematical theory used to describe interactions of elementary particles, the Standard Model (SM). The elementary particles and their interactions were first discussed and the consequences of their being three generations of fermions introduced. This allows mixing between the flavour eigenstates, in which the fermions interact, and the mass eigenstates, in which they propagate. The level of mixing is characterised by the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which also allows for CP-violation in interactions that involve transitions between quark generations. This manifests as CP-violation in decays of mesons and baryons.

The different types of CP-violation that can occur were then discussed. Direct CP-violation occurs if the amplitudes of a decay and its CP conjugate decay differ in that $\left|A_f/\bar{A}_{\bar{f}}\right| \neq 1$. The SM predicts that CP-violation in interactions involving charm (c) quarks is $\mathcal{O}(10^{-3})$ or less. It is thus very exciting that a recent LHCb result has measured direct CP-violation at $\mathcal{O}(10^{-3})$ in decays of the D^0 meson [1], which consists of $c\bar{\mathrm{u}}$ valence quarks. Mixing also occurs in systems of neutral mesons, such as the D^0, whereby the D^0 transforms itself into a $\overline{\mathrm{D}}^0$, and vice versa. Indirect CP-violation in mixing occurs if $|q/p| \neq 1$, where q and p are the coefficients of the flavour eigenstates of the D^0 in the definition of the mass eigenstates. An additional form of CP-violation can occur if the final state is accessible to both the meson and anti-meson. In this case the decays of un-mixed and mixed states can interfere and cause indirect CP-violation, even in the case that CP-violation is conserved in both mixing and decay.

The parameters y_{CP} and A_Γ were then introduced. y_{CP} examines the difference between the average decay rate of the D^0 and $\overline{\mathrm{D}}^0$ to a CP-eigenstate, to its average decay rate to a

CP-undefined final state:

$$y_{CP} = \frac{\hat{\Gamma}(\mathrm{D}^0 \to f) + \hat{\Gamma}(\overline{\mathrm{D}}^0 \to f)}{2\Gamma_{\mathrm{D}^0}} - 1$$
$$= \frac{\tau_{\text{eff}}(\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{\mp}\pi^{\pm})}{\tau_{\text{eff}}(\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{+}\mathrm{K}^{-})} - 1, \tag{7.1}$$

where τ_{eff} is the average proper decay time, or 'effective lifetime', of the $\overset{(-)}{\mathrm{D}}{}^0$ in the decay. $\mathrm{K}^{\mp}\pi^{\pm}$ is chosen as the CP-undefined final state as it is Cabibbo favoured, and so benefits from a large branching fraction. $\mathrm{K}^{+}\mathrm{K}^{-}$ is the CP-even final state with the largest branching fraction. As y_{CP} is calculated using the combined effective lifetime of the D^0 and $\overline{\mathrm{D}}^0$ the flavour of the $\overset{(-)}{\mathrm{D}}{}^0$ at production need not be known. A_{Γ} examines the difference in the average decay rate between an initial state of D^0 and $\overline{\mathrm{D}}^0$ decaying to a CP-eigenstate:

$$A_{\Gamma} = \frac{\hat{\Gamma}(\mathrm{D}^0 \to f) - \hat{\Gamma}(\overline{\mathrm{D}}^0 \to f)}{\hat{\Gamma}(\mathrm{D}^0 \to f) + \hat{\Gamma}(\overline{\mathrm{D}}^0 \to f)}$$
$$= \frac{\tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^{+}\mathrm{K}^{-}) - \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^{+}\mathrm{K}^{-})}{\tau_{\text{eff}}(\overline{\mathrm{D}}^0 \to \mathrm{K}^{+}\mathrm{K}^{-}) + \tau_{\text{eff}}(\mathrm{D}^0 \to \mathrm{K}^{+}\mathrm{K}^{-})}. \tag{7.2}$$

Here the flavour of the $\overset{(-)}{\mathrm{D}}{}^0$ must be known at production, and so the chain $\mathrm{D}^{*\pm} \to \overset{(-)}{\mathrm{D}}{}^0\pi^{\pm}$ is used. The charge of the $\pi^{\pm}$ thus determines the flavour of the $\overset{(-)}{\mathrm{D}}{}^0$. In the absence of CP-violation y_{CP} will be measured to be consistent with the mixing parameter $y = \Delta\Gamma/\Gamma_{\mathrm{D}^0}$, and A_{Γ} will be consistent with zero. New, non-SM, particles can potentially enter into the mixing and decay processes and enhance the level of CP-violation, thus increasing the size of A_{Γ} and the deviation of y_{CP} from y. The current world best measurements of y_{CP} and A_{Γ} were made by BABAR and BELLE, and show no evidence for indirect CP-violation.

Chapter 2 then discussed the experimental setup of the LHCb detector at the LHC. Each of the sub-detectors were discussed in turn, and their excellent performance during the 2010 data-taking run presented. The trigger systems used to decide which events to keep and which to discard were also detailed, as well as the offline data processing required for any physics analyses to be performed. LHCb is well designed for measuring the lifetime of the D^0, as is required for y_{CP} and A_{Γ}. The Vertex Locator (VELO) provides precise measurements of the positions of the proton-proton collisions within LHCb (primary vertices or PVs) as well as any displaced vertices produced by the decay of long lived particles. It thus provides a proper-decay-time resolution of $\sim$50 fs, which is much smaller than the average lifetime of the D^0. The Ring Imaging Cherenkov Detectors (RICH) provide very accurate particle identification, allowing clean separation of πs and Ks. This is essential in distinguishing the $\mathrm{K}^{\mp}\pi^{\pm}$ and $\mathrm{K}^{+}\mathrm{K}^{-}$ final states. Finally, the production cross section of D^0 at LHCb is very large, allowing large numbers of D^0 decays to be reconstructed.

The performance of the VELO was further evaluated in chapter 3, which examined the resolutions it achieves on impact parameter (IP) measurements. IP measurements are essential in the trigger, as the decay products of long lived particles like the D^0 tend to have larger IPs. The resolution with which IPs can be determined is also reflected in the vertex and proper-decay-time resolutions. A parametrisation of IP resolutions was derived, depending on the single hit resolution in the VELO, the material budget, and the extrapolation distance to the interaction point. A method of measuring IP resolutions without the use of simulation was then introduced, and its results compared to this parametrisation. In general IP resolutions were found to vary as expected with the various parameters involved. The performance of the VELO in this respect was also found to be excellent, achieving a resolution on IP_x of $< 36\ \mu\mathrm{m}$ for particles with $p_T > 1\,\mathrm{GeV}$. A detailed comparison between 2011 data and simulated data was made, and both were compared to the predictions of the mathematical parametrisation. A momentum dependent discrepancy between 2011 and simulated data was observed, whereby the IP resolutions on 2011 data are $\sim$20 % worse at low momentum than on simulated data. This effect is confined to the regions of the VELO away from that in which the two halves of the VELO overlap; within the overlap region the resolutions on 2011 and simulated data agree well. As this discrepancy is momentum dependent it suggests an issue with the description of the VELO material or the modelling of multiple scattering in material in the simulation. Complementary studies have not found any major issues in either of these areas, though the shape of the RF-foil, which encases the VELO, is known to be simplified in the simulation. Studies are underway to determine the effects of using a more accurate description of the RF-foil in the simulation. Thus, the exact source of the discrepancy between IP resolutions on data and simulation remains to be found.

The methods used to extract the effective lifetime of a specific decay from a dataset containing signal as well as different backgrounds were discussed in chapter 4. First, the general methods of extracting the optimal parameters of probability density functions (PDFs) from data using maximum likelihood fits was detailed. Also discussed was how this can be used to statistically distinguish signal and backgrounds, and determine their optimal descriptions and relative fractions in the dataset. For flavour tagged D^0 one can use a simultaneous fit to the distributions of $m(\mathrm{D}^0)$ and $\Delta m \equiv m(\mathrm{D}^{*\pm}) - m(\mathrm{D}^0)$ to distinguish signal, randomly-tagged D^0, combinatorial backgrounds and potentially any three body backgrounds that may need to be taken into account in future. Secondary D^0, produced in $\mathrm{B} \to \mathrm{D}^0 X$ decays, cannot be distinguished using the distributions of $m(\mathrm{D}^0)$ or Δm. However, as the B is not reconstructed they can be distinguished using the $\chi^2(IP_{\mathrm{D}^0})$, which tends to take larger values at high reconstructed proper decay times than prompt D^0. Thus, secondary D^0 are distinguished using a simultaneous fit to the proper-decay-time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distributions.

The method by which the effective lifetime of the signal is determined was then covered. Firstly, the proper-decay-time PDFs that can be used to model prompt and secondary D^0, taking into account the effects of non-zero proper-decay-time resolution, were presented. This was followed by a description of the data-driven method of correcting for lifetime biasing candidate selection criteria, the 'swimming algorithm', which is key to the measurements presented here. This technique exploits the fact that the kinematics of the decay products of a D^0 are independent of the proper decay time of the D^0 itself. This allows one to artificially change the proper decay time of the D^0 by moving any PVs in the event in the direction of the D^0 momentum. The decision of the candidate selection is then re-evaluated at each proper decay time. Thus, one can calculate the selection efficiency as a function of proper decay time for each D^0 candidate. The fact that the High Level Trigger (HLT) at LHCb is implemented in software is also key to this method, as it allows the trigger to be re-run identically as was done during data-taking. The technicalities of including the per-candidate acceptance functions calculated by this algorithm in the fit PDF were also discussed. Finally, the general form of the full fit PDF was presented. This PDF provides full discrimination between signal and all backgrounds, accounts for detector resolution, and corrects for lifetime biasing candidate selection criteria. It can thus be used to extract the effective lifetime of the signal.

Chapter 5 then presented the measurements of y_{CP} and A_Γ made using the techniques presented in chapter 4. The data used were collected by LHCb during the 2010 data-taking run, and comprise $28.0 \pm 2.8\,\mathrm{pb}^{-1}$. The specific trigger and offline selections applied to the data were detailed. The final datasets comprise 286,159 $\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^\mp\pi^\pm$ and 39,263 $\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^+\mathrm{K}^-$ candidates.

The results of fits to the distributions of $m(\mathrm{D}^0)$ and Δm to determine the fractions of signal, randomly-tagged D^0, and combinatorial backgrounds were then shown. As the level of combinatorial backgrounds is very low, the fraction of combinatorial $\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^+\mathrm{K}^-$ candidates cannot be determined accurately. Thus, only results for $\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^\mp\pi^\pm$ were shown. These found that the datasets consist of $\sim$99.2 % signal, of which $\sim$95.8 % has the $\mathrm{D}^{*\pm}$ correctly reconstructed.

Following this, the specifics of the fits to extract effective lifetimes were discussed. As the level of combinatorial backgrounds is so low, their effect is neglected in these fits and a corresponding systematic uncertainty assigned. The fits to the proper-decay-time and $\ln(\chi^2(IP_{\mathrm{D}^0}))$ distributions thus consider only prompt and secondary D^0. The bias resulting from mis-tagged D^0 is accounted for after the determination of the effective lifetimes, using the random-tag rates determined by the mass fits to the $\overset{(-)}{\mathrm{D}^0} \to \mathrm{K}^\mp\pi^\pm$ datasets. The

specific PDFs used to describe the data were then presented, as well as the manner in which the acceptance variables determined by the swimming algorithm are accounted for in the fit. As an accurate parametrisation of secondary D^0 is difficult to achieve with the data available an additional fit iteration is performed on a subset of the data in which the fraction of secondary D^0 is suppressed. This reduces the final datasets to 226,110 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{\mp}\pi^{\pm}$ and 30,481 $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{+}\mathrm{K}^{-}$. The results of the fits using these PDFs and the effective lifetimes and their statistical uncertainties thus determined on each dataset were then given. The quality of these fits was assessed and found to be sufficiently good. The values of y_{CP} and A_{Γ}, and their statistical uncertainties, were then shown.

Chapter 6 presented checks of the stability of the measurements of y_{CP} and A_{Γ}, and the evaluation of their systematic uncertainties. The datasets were split into several subsets to evaluate any dependencies in the results. The values of y_{CP} and A_{Γ} obtained on all subsets of the data were found to be consistent within their statistical uncertainties, showing them to be stable. A cross-check measurement of $A_{\Gamma}^{\mathrm{K}\pi,\mathit{eff}}$ was also made. This is defined analogously to A_{Γ}, but using $\tau_{\mathrm{eff}}(\mathrm{D}^0 \to \mathrm{K}^{-}\pi^{+})$ and $\tau_{\mathrm{eff}}(\overline{\mathrm{D}}{}^0 \to \mathrm{K}^{+}\pi^{-})$, and so exploits the high statistics of the $\overset{(-)}{\mathrm{D}}{}^0 \to \mathrm{K}^{\mp}\pi^{\pm}$ channel. It was found to be consistent with zero, as expected, again showing the results to be reliable. The results of many pseudo experiments on toy Monte Carlo simulated data were then shown. The only significant bias found was determined to be due to neglecting combinatorial backgrounds in the fit.

Finally, various sources of systematic uncertainty in the results were considered. The dominant systematics were found to result from neglecting combinatorial backgrounds and the parametrisation of secondary D^0. These result in the final measurements of A_{Γ} and y_{CP}

$$A_{\Gamma} = \quad (-5.9 \pm 5.9\,(\mathrm{stat.}) \pm 2.1\,(\mathrm{syst.})) \times 10^{-3}, \tag{7.3a}$$

$$y_{CP} = \quad (5.5 \pm 6.3\,(\mathrm{stat.}) \pm 4.1\,(\mathrm{syst.})) \times 10^{-3}. \tag{7.3b}$$

A_{Γ} is consistent with zero and y_{CP} is consistent with the world average measurement of $y = (8.0 \pm 1.3) \times 10^{-3}$ [3] and with zero. Thus, these results show no evidence for *CP*-violation or mixing. Although the statistical uncertainties attained are not yet competitive with previous measurements made at the B factories they make an important contribution to the world average.

7.2 Outlook

7.2.1 Current and Near Future LHCb Data

The dataset already collected in 2011 comprises 1107 pb^{-1} , which is $\sim$39 times as large as that used for the analyses presented here. This could reduce the statistical uncertainty on y_{CP} and A_Γ by a factor of $\sim$6, and so has the potential to achieve a precision of 1×10^{-3}. This would of course require controlling the systematic uncertainties to this level. Should this be achieved these would represent the world best measurements by far, and the first 5σ single measurement of D^0 mixing, via y_{CP}. Given the observation by LHCb of direct CP-violation in the D^0 system at $\mathcal{O}(10^{-3})$, this provides real promise for observing indirect CP-violation as well.

The largest contributing factors to the systematic uncertainties on the measurements presented here are from the parametrisation of secondary D^0 and neglecting combinatorial backgrounds in the fit. Improvements in the trigger selections mean that the 2011 datasets allow access to much larger, clean samples of both these backgrounds. Thus, using 2011 data and possibly some of the methods detailed in section 4.5 these systematic effects should be greatly reduced from the values presented here.

The 2012 run should provide at least as much integrated luminosity as in 2011. The data will also be taken at $\sqrt{s} = 8$ TeV, causing an increase in the D^0 production cross section. The available trigger and permanent storage capacity for charm physics has also been significantly increased. This will extend the reach of these measurements below 10^{-3} and reach a sensitivity at which the effects of new physics beyond the SM could become apparent. The measurement of $\Delta\mathcal{A}^{\mathrm{CP}}$, which provides the first evidence for direct CP-violation in the D^0 system, will reach similar precision. This could potentially resolve the debate as to whether its measured value is allowed within the SM, or if new physics is at play. These datasets will also yield a significant sample of 'wrong sign' $\mathrm{D}^0 \to \mathrm{K}^+\pi^-$ decays. These give access to the mixing parameters x'^2 and y'^2, which are related to x and y by a strong phase. A time dependent Dalitz analysis of $\mathrm{D}^0 \to \mathrm{K}^0_{\mathrm{S}} h^+ h^-$ will also give access to x and y. These should be able to obtain sensitivities of $\mathcal{O}(10^{-3})$. CP-violation is also being searched for in other three and four body decay modes.

Figure 7.1 shows the projection to 2017 of the measurements of $\Delta\mathcal{A}^{\mathrm{CP}}$ and A_Γ performed by LHCb, assuming the same central values are obtained. In order to achieve sensitivities at or below $\mathcal{O}(10^{-4})$ the high luminosity regime of the LHCb upgrade would be required.

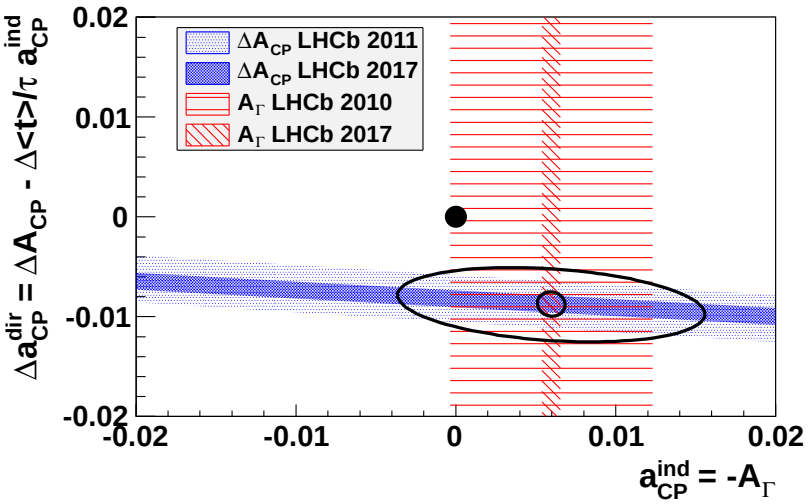


Figure 7.1: The contour plot of direct vs. indirect CP-violation in the D^0 system using the current values of the $\Delta\mathcal{A}^{\mathrm{CP}}$ and A_Γ measurements performed at LHCb. The inner contour shows the projection of these values to 2017, assuming the same central values and a reduction in the total uncertainties by a factor of $1/\sqrt{N}$. Reproduced from [100].

7.2.2 The LHCb Upgrade

LHCb is fast becoming a world leader in flavour physics and has already collected some of the largest datasets yet recorded for many key decay channels. However, the recorded integrated luminosity, and thus the size of these datasets, will scale linearly with running time, while the statistical precision that can be attained scales as $1/\sqrt{N}$. For example, after five years of nominal operation an additional five years of data-taking would only improve the precision achieved by a factor of $1/\sqrt{2}$. With any luck five years will be sufficient for LHCb to discover many indications of physics beyond the SM that will require probing at much higher precision. Thus, an upgrade is required to increase the instantaneous luminosity at which LHCb operates, and so increase the rate at which this precision improves.

While LHCb has performed admirably at an instantaneous luminosity of up to $\sim 4 \times 10^{32}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ the L0 trigger becomes very inefficient for hadronic decay modes at higher luminosities. This is because it only has access to information from the calorimeters and muon stations, which can be read out at 40 MHz. This information is insufficient to efficiently trigger decays of long lived particles at high luminosity while keeping retention rates low enough to fit within the timing constraints of the trigger. Thus, the upgraded detector is intended to be able to readout information from all its sub-detectors at 40 MHz [101]. This would allow the first level trigger to perform more complete event reconstruction, and trigger on the presence of displaced decay vertices. This both greatly increases the efficiency of the first level trigger for hadronic decays, and allows the upgraded detector to potentially operate at instantaneous

luminosities up to $\sim 2 \times 10^{33}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. At this rate the upgraded detector would collect in one year of data-taking what the current detector will collect in five to ten years. The upgraded detector is currently at the design stage, with the intention of beginning installation in 2017. The operational instantaneous luminosity of the LHCb upgrade is still lower that the maximum design luminosity of the LHC at present. Hence, the LHCb upgrade is not contingent on an upgrade to the LHC, but is compatible with the future high luminosity LHCb running phase.

For high yield decay channels like $\mathrm{D}^0 \to \mathrm{K}^-\pi^+$ and $\mathrm{D}^0 \to \mathrm{K}^+\mathrm{K}^-$ the LHCb upgrade will afford the opportunity to measure indirect CP-violation in the D^0 system, via y_{CP} and A_Γ, to a precision below $\mathcal{O}(10^{-4})$. Direct CP-violation will also be able to reach similar levels of precision, via measurements like $\Delta\mathcal{A}^{\mathrm{CP}}$. The Dalitz analysis of the decay $\mathrm{D}^0 \to \mathrm{K}^0_\mathrm{S}h^+h^-$ and wrong sign $\mathrm{D}^0 \to \mathrm{K}^+\pi^-$ will also provide measurements of the mixing parameters x and y of a similar accuracy. Measurements at this precision will provide exceptionally strong tests of the SM and potentially insight into the nature of physics beyond the SM. The measurements made in the B sector using the datasets collected by an upgraded LHCb will also provide stringent tests on the predictions of the SM and the CKM mechanism. These include various complementary measurements via different decay channels of the CKM angle γ to a precision of less than 1°, to examine if the unitarity triangles are indeed unitary. A measurement at an accuracy of 1 % will also be possible on the 'zero crossing point' in the forward-backward asymmetry of the decay $\mathrm{B}^0 \to \mathrm{K}^{*0}\mu^+\mu^-$, which can be strongly influenced by new physics. Thus, the LHCb upgrade presents the possibility of a new era in high precision tests of the SM. Indeed, the additional flexibility introduced in the LHCb trigger in the upgrade will also allow the detector to become a more general purpose experiment in the forward region. This will extend its reach in areas such as lepton flavour violating τ decays, electroweak studies, and long lived exotics.

7.2.3 Other Flavour Physics Experiments

While LHCb and the LHCb upgrade provide exciting prospects for the future of flavour physics, LHCb is far from the only flavour physics experiment currently planned. Historically, flavour physics measurements have been performed at precision, asymmetric e^+-e^- colliders, which benefit from much lower backgrounds than hadronic machines like the LHC. Such experiments include BABAR on the PEPII collider at SLAC, and BELLE on the KEKB collider at KEK. The measurements of y_{CP} and A_Γ performed at BABAR and BELLE are currently the world bests. The D0 and CDF experiments at the TEVATRON p-$\overline{\mathrm{p}}$ collider have also made valuable contributions to flavour physics both in the D and B sectors. While all these experiments have stopped data-taking in recent years there are some exciting prospects

on the horizon for experiments to complement and challenge LHCb.

The KEKB collider is being upgraded to SuperKEKB [102], which intends to achieve a maximum instantaneous luminosity of $8 \times 10^{35}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This would allow an integrated luminosity of 50 ab^{-1} to be accumulated by 2022. This requires an upgrade of the detector, Belle-II, to deal with higher sensor occupancies and improve vertexing and PID performance. Another prospect for the future is the SuperB experiment, proposed to be built in at the Cabibbo Laboratory, in Italy [103]. It too is an asymmetric e^+-e^- collider that aims to operate initially at an instantaneous luminosity of $10^{36}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$, and collect 75 ab^{-1} within five years of data-taking. These B factories operate primarily with $\sqrt{s}$ equal to the mass of the $\Upsilon(4S)$ resonance, which decays to quantum correlated $\mathrm{B}\,\overline{\mathrm{B}}$ pairs. This will allow them to collect several tens of billions of such pairs. Large samples of D^0 decays will also be collected via $\mathrm{B} \to \mathrm{D}^0 X$ and $\mathrm{e}^+\mathrm{e}^- \to \mathrm{q}\overline{\mathrm{q}}$. They could potentially also operate for some time at the $\psi(3770)$ resonance, which produces $\mathrm{D}^0\ \overline{\mathrm{D}}^0$ pairs. As these pairs are quantum correlated they would offer sensitivities to phase differences in D^0 decays.

The LHCb upgrade will benefit from higher production cross sections than the future B factories, and will thus obtain the highest precision in channels in which all decay products are charged. However, due to the cleaner environment of the B factories they will be able to study channels with neutral decay products, which are very difficult to perform at LHCb. They will also be able to search for very rare SM decays, such as $\mathrm{B} \to \ell\nu$, in which LHCb cannot compete. The ability of SuperB to partially polarise its electron beams may also help to reduce backgrounds. Additionally, the fact that B and D mesons are produced in quantum correlated pairs allows determination of the strong phase difference in mixing, which will need to be used as input to the measurement of x and y via $\mathrm{D}^0 \to \mathrm{K}^0_\mathrm{S} h^+ h^-$. Thus, the measurements performed at an upgraded LHCb and the future B factories will both complement and compete with each other.

Thus, the coming years present many possibilities in testing the SM to its limits and beyond. These could lead to the discovery of new physics and provide a deeper understanding of the nature of the most basic elements of the universe. However the coming years play out, it is a very exciting time to be involved in flavour physics.

Bibliography

[1] The LHCb Collaboration. Evidence for CP violation in time-integrated $D^0 \to h^- h^+$ decay rates. *arXiv:1112.0938v1 [hep-ex]*, 2011. LHCb-PAPER-2011-023.

[2] The BaBar Collaboration. Measurement of D^0-$\bar{D}^0$ Mixing using the Ratio of Lifetimes for the Decays $D^0 \to K^-\pi^+$ and K^+K^-. *Phys. Rev. D*, 2009. 80:071103.

[3] The Heavy Flavor Averaging Group. `http://www.slac.stanford.edu/xorg/hfag/`, January 2012.

[4] The Belle Collaboration. Evidence for $D^0 - \bar{D}^0$ Mixing. *Phys. Rev. Lett.*, 2007. 98:211803.

[5] F. Halzen and A. D. Martin. *Quarks & Leptons*. John Wiley and Sons, 1984.

[6] Stephen P. Martin. A Supersymmetry Primer. arXiv:hep-ph/9709356v6, September 2011.

[7] K. Nakamura et al. Review of Particle Physics. *J. Phys. G.*, 37, 2011. 075021.

[8] W. Pauli. The Connection Between Spin and Statistics. *Phys. Rev.*, 58:716–722, 1940.

[9] Y. Fukuda et al. Evidence for Oscillation of Atmospheric Neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.

[10] The ATLAS Collaboration. Search for Quark Contact Interactions in Dijet Angular Distributions in pp Collisions at $\sqrt{s}$ = 7 TeV Measured with the ATLAS Detector. *Phys. Lett. B*, 694:327–345, 2010.

[11] V. E. Barnes et al. Observation of a Hyperon with Strangeness Number Three. *Physical Review Letters*, 12:204, 1964.

[12] S. Davidson. The Standard Model. In *Proceedings of the School for Experimental High Energy Physics Students, 1–14 September 2009, Oxford, UK*, pages 115–209, 2010. RAL-TR-2010-010.

[13] P. W. Higgs. Broken symmetries, massless particles and gauge fields. *Phys. Lett.*, 12:132–133, 1964.

[14] The ATLAS Collaboration. An update to the combined search for the Standard Model Higgs boson with the ATLAS detector at the LHC using up to 4.9 fb^{-1} of pp collision data at $\sqrt{s} = 7$ TeV. ATLAS-CONF-2012-019, 2012.

[15] The CMS Collaboration. Combined results of searches for a Higgs boson in the context of the standard model and beyond-standard models. CMS PAS HIG-12-008, 2012.

[16] S. L. Glashow. Partial Symmetries of Weak Interactions. *Nucl. Phys. B*, 22:579–588, 1961.

[17] A. Salam and J. C. Ward. Electromagnetic and weak interactions. *Phys. Lett.*, 13:168–171, 1964.

[18] S. Weinberg. A Model of Leptons. *Phys. Rev. Lett.*, 19:1264–1266, 1967.

[19] M. Kaku. *Quantum Field Theory*. Oxford University Press, 1993.

[20] N. Cabibbo. Unitary Symmetry and Leptonic Decays. *Phys. Rev. Lett.*, 10:532–533, 1963.

[21] S. L. Glashow, J. Iliopoulos, and L. Maiani. Weak Interactions with Lepton-Hadron Symmetry. *Phys. Rev. D*, 2:1285–1292, 1970.

[22] M. Kobayashi and T. Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Prog. Theor. Phys.*, 49:652–657, 1973.

[23] L. Wolfenstein. Parameterisation of the Kobayashi-Maskawa matrix. *Phys. Rev. Lett.*, 51:1945, 1983.

[24] V. Fanti et al (NA48 Collab.). A new measurement of direct CP violation in two pion decays of the neutral kaon. *Phys. Lett. B*, 465:335–348, 1999.

[25] A. Alavi-Harati et al (KTeV Collab.). Observation of Direct CP Violation in $K_{S,L} \to \pi\pi$ Decays. *Phys. Rev. Lett.*, 83:22–27, 1999.

[26] B. Aubert et al (Babar Collab.). Direct CP Violating asymmetry in $B_d \to K^+\pi^-$ decays. *Phys. Rev. Lett.*, 93(131801), 2004.

[27] Y. Chao et al (Belle Collab.). Evidence for direct CP violation in $B_d \to K^+\pi^-$ decays. *Phys. Rev. Lett.*, 93(191802), 2004.

[28] J. H. Cristenson et al. Evidence for the 2π decay of the K_2^0 meson. *Phys. Rev. Lett.*, 13:138–140, 1964.

[29] The LHCb Collaboration. Prompt charm production in pp collisions at $\sqrt{s}$ = 7 TeV. LHCb-CONF-2010-013, 2010.

[30] M. Gersabeck, M. Alexander, S. Borghi, V. V. Gligorov, C. Parkes. On the interplay of direct and indirect CP violation in the charm sector. *J. Phys. G: Nucl. Part. Phys.*, 39, 2012. 045005.

[31] The Heavy Flavour Averaging Group. Averages of b-hadron, c-hadron, and τ-lepton Properties. *arXiv:1010.1589 [hep-ex]*, 2011.

[32] The BaBar Collaboration. Measurement of D^0-$\bar{D}^0$ Mixing using the Ratio of Lifetimes for the Decays $D^0 \to K^-\pi^+$ and K^+K^-. *Phys. Rev. D*, 2008. 78:011105.

[33] The LHCb Collaboration. Measurement of mixing and CP violation parameters in two-body charm decays. LHCb-PAPER-2011-032 [arXiv:1112.4698], 2011.

[34] Lyndon Evans and Philip Bryant (editors). LHC Machine. *JINST*, 3, 2008. S08001.

[35] CERN: The European Organization for Nuclear Research. `http://www.cern.ch`, April 2012.

[36] The ATLAS Collaboration. Luminosity Public Results. `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResults`, April 2012.

[37] S. Chatrchyan et al. The CMS experiment at the CERN LHC. *JINST*, 3(S08004), 2008.

[38] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3(S08003), 2008.

[39] K. Aamodt et al. The ALICE experiment at the CERN LHC. *JINST*, 3(S08002), 2008.

[40] A. Augusto Alves Jr et al. The LHCb Detector at the LHC. *JINST*, 3(S08005), 2008.

[41] LHCb Technical Design Report: Vertex Locator. CERN-LHCC-2001-011.

[42] LHCb Technical Proposal. CERN-LHCC-98-004.

[43] OP Vistars. LHC Status. http://op-webtools.web.cern.ch/op-webtools/vistar/vistars.php, April 2012.

[44] T. Szumlak & C. Parkes. VELO Event Model. LHCb-2006-054.

[45] E. Belau et al. Charge Collection in Silicon Strip Detectors. *Nuclear Instruments and Methods*, 214:253–260, 1983.

[46] The Moore Project. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/moore/, July 2012.

[47] The Brunel Project. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/brunel/, July 2012.

[48] LHCb Technical Design Report: Trigger System. CERN-LHCC-2003-031.

[49] R. van der Eijk et al. Track reconstruction for LHCb. CERN-LHCb-98-045.

[50] VELO Material for Publications. https://lbtwiki.cern.ch/bin/view/VELO/VELOConferencePlots, April 2012.

[51] M. Krasowski et al. Primary vertex reconstruction. CERN-LHCb-2007-011.

[52] J. D. Jackson. *Classical Electrodynamics*. John Wiley and Sons, 2nd edition, 1975.

[53] N. Brook et al. LHCb RICH 1 Engineering Design Review Report. CERN-LHCb-2004-121, October 2005.

[54] M. Adinolfi et al. LHCb RICH 2 engineering design review report. CERN-LHCb-2002-009, March 2002.

[55] A. MacGregor. *The Laser Mirror Alignment System for the LHCb RICH Detectors*. PhD thesis, University of Glasgow, April 2006.

[56] LHCb Technical Design Report: RICH. CERN-LHCC-2000-037.

[57] LHCb: From the Detector to the first Physics Results - HEP-MAD Conference. http://www.lpta.univ-montp2.fr/users/qcd/hep.html, August 2011.

[58] LHCb Technical Design Report: Magnet. CERN-LHCC-2000-007.

[59] Dr Fred Blanc. Private correspondence, 2012.

[60] J. Gassner et al. Layout and Expected Performance of the LHCb TT Station. CERN-LHCb-2003-140.

[61] ST Material for Publications. http://lhcb.physik.uzh.ch/ST/public/material/index.php, August 2011.

[62] LHCb Technical Design Report: Reoptimized Detector Design and Performance. CERN-LHCC-2003-030.

[63] LHCb Technical Design Report: Outer Tracker. CERN-LHCC-2001-024.

[64] A high-performance tracking system for the LHCb spectrometer. https://indico.cern.ch/conferenceDisplay.py?confId=149003, September 2011.

[65] LHCb Technical Design Report: Calorimeters. CERN-LHCC-2000-036.

[66] LHCb Technical Design Report: Muon System. CERN-LHCC-2001-010.

[67] A. Powell for the LHCb Collaboration. Particle Identification at LHCb. LHCb-PROC-2011-008, 2011.

[68] The DaVinci Project. http://lhcb-release-area.web.cern.ch/LHCb-release-area/DOC/davinci/, July 2012.

[69] GEANT4: A toolkit for the simulation of the passage of particles through matter. http://geant4.cern.ch/.

[70] PYTHIA. http://home.thep.lu.se/~torbjorn/Pythia.html.

[71] The EvtGen package. http://www.slac.stanford.edu/~lange/EvtGen/.

[72] A. Papadelis. *Characterisation and commissioning of the LHCb VELO detector.* PhD thesis, Vrije Universiteit, June 2009. CERN-THESIS-2009-044.

[73] G. R. Lynch and O. I. Dahl. Approximations to multiple Coulomb scattering. *Nucl. Instrum. Methods*, 1991. 6.

[74] M. Whitehead. Vertex Locator Simulated Material Description. CERN-LHCb-INT-2010-054, 2010.

[75] The LHCb VELO Group. Performance of the LHCb Vertex Locator. CERN-PH-EP-2012-XXX, 2011. To be submitted to JINST.

[76] M. Reid, P. Collins, T. Gershon. Geant4 Impact Parameter Resolution Study Leptons - e^- & μ^-. LHCb-INT-2011-034, 2011.

[77] M. Reid, P. Collins, T. Gershon. Geant4 Impact Parameter Resolution Study Hadrons - π^-. LHCb-INT-2011-036, 2011.

[78] S. Viret, C. Parkes, D. Petrie. LHCb VELO software alignment Part I: the alignment of the VELO modules in their half boxes. CERN-LHCb-2005-101, 2005.

[79] S. Viret, C. Parkes, M. Gersabeck. LHCb VELO software alignment Part II: the alignment of the VELO detector-halves. CERN-LHCb-2007-067, 2007.

[80] M. Gersabeck, C. Parkes, S. Viret. LHCb VELO software alignment Part III: the alignment of the relative sensor positions. CERN-LHCb-2007-138, 2007.

[81] MINUIT : Function Minimization and Error Analysis. `http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html`.

[82] Giovanni Punzi. Comments on Likelihood fits with variable resolution. arXiv:physics/0401045v1 [physics.data-an], 2004.

[83] K. Cranmer. Kernel Estimation in High Energy Physics. *Comput. Phys. Commun.*, 136:198–207, 2001.

[84] M. Gersabeck. *Alignment of the LHCb Vertex Locator and Lifetime Measurements of Two-Body Hadronic Final States.* PhD thesis, University of Glasgow, September 2009.

[85] The LHCb Collaboration. Measurement of the effective $B_s^0 \to K^+K^-$ lifetime. *Phys. Lett. B*, D11, 2011. 01604R1.

[86] J. E. Gaiser. *Charmonium Spectroscopy from Radiative Decays of the J/Psi and Psi-Prime - Appendix F.* PhD thesis, SLAC, 1982.

[87] M. Alexander *et al*. Time-Dependent Analyses of Two-Body Charm Decays. LHCb-ANA-2011-066, 2011.

[88] The LHCb Collaboration. Measurement of J/ψ production in pp collisions at $\sqrt{s}$ = 7 tev. *Eur. Phys. J.*, C71:1645, 2011.

[89] Y. Xie. sFit: a method for background subtraction in maximum likelihood fit. arXiv:0905.0724v1, 2009.

[90] R. Bailey *et al.*. Measurement of the lifetime of charged and neutral d mesons with high resolution silicon strip detectors. *Z. Phys.*, pages 357–363, 1885.

[91] The DELPHI Collaboration. Lifetimes of charged and neutral b hadrons using event topology. *Z. Phys.*, pages 363–374, 1995.

[92] J. Rademacker. Reduction of the statistical power per event due to upper lifetime cuts in lifetime measurements. *Nucl. Instrum. Meth*, A570:525–528, 2007. [arXiv:hep-ex/0502042v3].

[93] F. Azfar et al. A Monte Carlo Independent Method for Lifetime Fits in Data biased by the Hadronic Trigger. CDF-6756, November 2003.

[94] The CDF Collaboration. Measurement of the B^- lifetime using a simulation free approach for trigger bias correction. *Phys. Rev. D*, 83, 2011. 032008.

[95] V. V. Gligorov. *Measurement of the CKM angle gamma and B meson lifetimes at the LHCb detector*. PhD thesis, University of Oxford, 2008. CERN-THESIS-2008-044.

[96] Muriel Pivk and Francois R. Le Diberder. sPlot: a statistical tool to unfold data distributions. *Nucl. Instrum. Meth. A*, 555:356–369, 2005.

[97] Dr. Silvia Borghi. Studies of Δm. `https://indico.cern.ch/getFile.py/access?contribId=4&resId=0&materialId=slides&confId=119902`, 2011.

[98] The RooFit Toolkit for Data Modeling. `http://roofit.sourceforge.net/`.

[99] C. Parkes, S. Borghi. Systematics on lifetime measurments. `https://lbtwiki.cern.ch/bin/view/VELO/VELOPlotsAndPublications`, 2012.

[100] Dr. Marco Gersabeck. Private correspondence, 2012.

[101] The LHCb Collaboration. Letter of Intent for the LHCb Upgrade. CERN-LHCC-2011-001, 2011.

[102] The Belle II Collaboration. Status and prospects of SuperKEKB collider and Belle II experiment. arXiv:1201.1248v2 [hep-ex], 2012.

[103] Adrian Bevan. The Physics Programme at SuperB. arXiv:1112.1394v1 [hep-ex], 2011.